Suppose we have an experiment and a sample space Ω.
We know want to introduce a <u>probabitity law</u>.

- which assigns a likelihood to any event in Ω

  if $A \subseteq \Omega$, then $P(A)$ is a number that tells us how likely the event A is (to occur).

However this assignment is done, it <u>MUST</u> always satisfy the following <u>**axioms**</u>

$\boxed{1.}$ (Non-negativity) $P(A) \geq 0$ for every event $A \subseteq \Omega$

$\boxed{2.}$ (Additivity) If $A, B \subseteq \Omega$ are disjoint, then

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

and, more generally, if $A_1, A_2, \ldots$ is any sequence (finite or infinite) of <u>disjoint</u> events, then

$$P(A_1 \cup A_2 \cup \cdots) = P(A_1) + P(A_2) + \cdots$$

$\boxed{3.}$ (Normalization) $P(\Omega) = 1$.

These **axioms** conform to our intuition.
For instance $\boxed{3}$ says Ω contains all possible outcomes.

These **axioms** also imply many properties that EVERY probability law satisfies.

Let's demonstrate (develop) some of these properties

Ω, $\emptyset$ are events.

$\Omega \cap \emptyset = \emptyset$ so Ω and $\emptyset$ are disjoint

Also, $\Omega \cup \emptyset = \Omega$.

$$1 \overset{\boxed{3}}{=} P(\Omega) = P(\Omega \cup \emptyset) \overset{\boxed{2}}{=} P(\Omega) + P(\emptyset) \overset{\boxed{3}}{=} 1 + P(\emptyset).$$

So $\boxed{P(\emptyset) = 0}$

More generally, (and similarly)

if $A \subseteq \Omega$ is any event

then $A \cup A^c = \Omega$

and $A \cap A^c = \emptyset$ so A and A^c are disjoint

So

$$1 = P(\Omega) = P(A \cup A^c) = P(A) + P(A^c)$$

$\boxed{P(A) = 1 - P(A^c)}$ also $\boxed{P(A^c) = 1 - P(A)}$

These are the Complementary rules.

$\boxed{\text{We will develop more properties later}}$

Let's now develop our first probability law.

Let's first consider the case of a <u>discrete</u> sample space.

For example, when Ω_F is finite we can write

$$\Omega_F = \{\omega_1, \omega_2, \omega_3, \ldots, \omega_N\}.$$

This set has N (distinct) elementary outcomes $|\Omega_F| = N$.

NOTATION When A is a set we write $|A| = n$ to mean that the number of elements in the set A is n.

Mathematicians say "the Cardinality of A is n".

Some discrete sample spaces can be countably infinite say, $\Omega_I = \{\omega_1, \omega_2, \omega_3, \ldots\}$ and $|\Omega_I| = \infty$

Let's suppose we know (or have assigned) the values of $P(\{\omega_i\})$ for each $\omega_i \in \Omega$.

Since $\Omega = \{\omega_1\} \cup \{\omega_2\} \cup \{\omega_3\} \cup \ldots$ (i.e., Ω is the disjoint union of its elementary outcomes)

by Axiom $\boxed{2}$

$$P(\Omega) = 1 = P(\{\omega_1\}) + P(\{\omega_2\}) + \cdots$$

Also, since any event A can be written as a disjoint union of its members

$$P(A) = \sum_{\omega \in A} P(\{\omega\})$$

to compute $P(A)$ we simply <u>add</u> all the prob. masses $P(\{\omega\})$ from each ω that belongs to A.

So, for example

if $A = \{\omega_2, \omega_6, \omega_7, \omega_9, \omega_{12}\}$ then

$$P(A) = P(\{\omega_2\}) + P(\{\omega_6\}) + P(\{\omega_7\}) + P(\{\omega_9\}) + P(\{\omega_{12}\})$$

prob. mass at $\{\omega_i\}$

That is, if we know $P(\{\omega_i\})$ for all i

to compute $P(A)$ for any event A we just simply

<u>add</u> the probability masses for each outcome in A.

Shorthand: I may write $P(\{\omega_i\})$ as $P(\omega_i)$

But we must realize that a probability law has a set as it's input.

For <u>finite</u> Ω the Most important special case is

the case of <u>equally-likely outcomes</u>

in this case the elementary outcomes ω_i are equally-likely to occur, i.e., $P(\omega_i) = c$ are all equal.

Since $1 = P(\omega_1) + P(\omega_2) + \cdots + P(\omega_N) = Nc$

we have $c = \frac{1}{N}$ is the (equal) probability for each outcome of such an experiment.

So, in the case of equally-likely outcomes in Ω,

$$P(A) = \frac{|A|}{|\Omega|} \quad \text{for any event } A \subseteq \Omega.$$

This is because

$$P(A) = \sum_{\omega \in A} P(\omega) = \sum_{\omega \in A} \frac{1}{|\Omega|} = |A| \cdot \frac{1}{|\Omega|} = \frac{|A|}{|\Omega|}.$$

Experiments that can be modeled reasonably well as equally-likely outcomes

1. Rolling a balanced die n times in succession
2. Tossing a balanced coin n times in succession
3. Picking a set of five cards from a well-shuffled deck
4. Selecting m people from a group of n qualified candidates.

Example Suppose we toss a balanced 4-sided die twice.

(1,1)	(2,1)	(3,1)	(4,1)
(1,2)	(2,2)	(3,2)	(4,2)
(1,3)	(2,3)	(3,3)	(4,3)
(1,4)	(2,4)	(3,4)	(4,4)

Ω

$|\Omega| = 16$

Let A = a '3' is rolled

$|A| = 7$

So, $P(A) = \frac{7}{16}$.

DeMorgan's Laws

$$(A_1 \cup A_2)^c = A_1^c \cap A_2^c$$

and

$$(A_1 \cap A_2)^c = A_1^c \cup A_2^c .$$

The law holds for any number of sets. In particular, it holds for any finite number of events. A homework problems asks you to verify DeMorgan's laws using Venn diagrams. Let's assume this has been done. Then

$$(A_1 \cup A_2 \cup A_3)^c = (A_1 \cup A_2)^c \cap A_3^c = A_1^c \cap A_2^c \cap A_3^c .$$

also

$$(A_1 \cap A_2 \cap A_3)^c = (A_1 \cap A_2)^c \cup A_3^c = A_1^c \cup A_2^c \cup A_3^c ,$$

shows that DeMorgan's laws must hold for any 3 events.

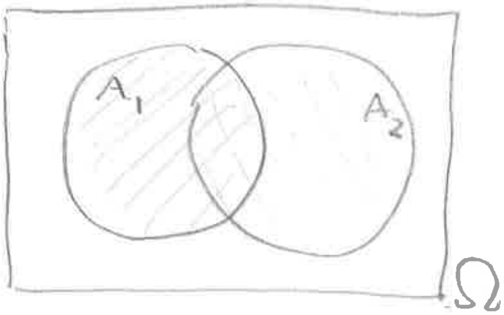
It is, in fact, true that if $A_1, A_2, \ldots, A_n$ are events, then

$$\left(\bigcup_{i=1}^{n} A_i \right)^c = \bigcap_{i=1}^{n} A_i^c \quad \text{and} \quad \left(\bigcap_{i=1}^{n} A_i \right)^c = \bigcup_{i=1}^{n} A_i^c .$$

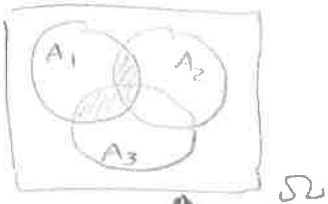
Remark By Complementarity, for example, we have

$$P\left(\bigcup_{i=1}^{n} A_i \right) = 1 - P\left(\left\{ \bigcup_{i=1}^{n} A_i \right\}^c \right) = 1 - P\left(\bigcap_{i=1}^{n} A_i^c \right).$$

Inclusion-exclusion principles (see §1.2 problems #12)



$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$$



$$P(A_1 \cup [A_2 \cup A_3]) = P(A_1) + P(A_2 \cup A_3) - P(A_1 \cap [A_2 \cup A_3])$$

$$= P(A_1) + \{P(A_2) + P(A_3) - P(A_2 \cap A_3)\} - P((A_1 \cap A_2) \cup (A_1 \cap A_3))$$

$$= P(A_1) + P(A_2) + P(A_3) - P(A_2 \cap A_3)$$

$$- \{P(A_1 \cap A_2) + P(A_1 \cap A_3) - P(\overbrace{A_1 \cap A_2 \cap A_1 \cap A_3}^{\text{same as } A_1 \cap A_2 \cap A_3})\}$$

$$\boxed{P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3) - P(A_1 \cap A_2) - P(A_1 \cap A_3) - P(A_2 \cap A_3) + P(A_1 \cap A_2 \cap A_3)}$$

this is the inclusion-exclusion principle for 3 sets.

We can, in fact, generalize this statement to any finite number of events. Here is the inclusion-exclusion for 4 events:

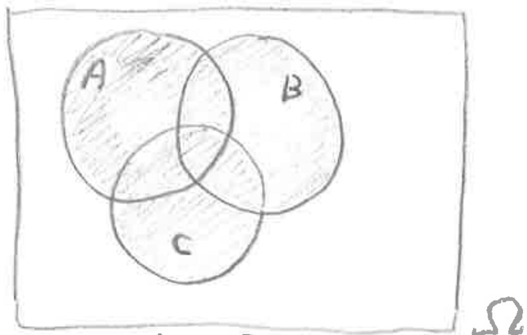
$$P(A_1 \cup A_2 \cup A_3 \cup A_4) = P(A_1) + P(A_2) + P(A_3) + P(A_4)$$

$$- P(A_1 \cap A_2) - P(A_1 \cap A_3) - P(A_1 \cap A_4) - P(A_2 \cap A_3) - P(A_2 \cap A_4) - P(A_3 \cap A_4)$$

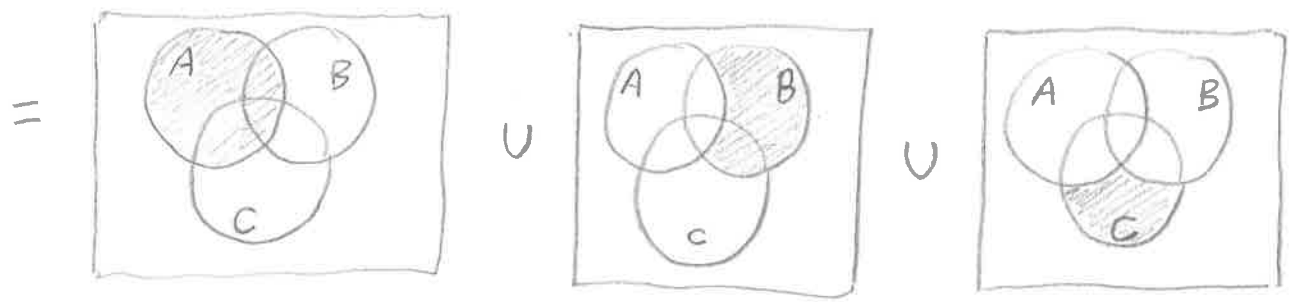
$$+ P(A_1 \cap A_2 \cap A_3) + P(A_1 \cap A_2 \cap A_4) + P(A_1 \cap A_3 \cap A_4) + P(A_2 \cap A_3 \cap A_4)$$

$$- P(A_1 \cap A_2 \cap A_3 \cap A_4).$$

Yet another representation for the probability of a union of events.



$A \cup B \cup C$



A $\qquad$ $A^c \cap B$ $\qquad$ $A^c \cap B^c \cap C$

So

$$P(A \cup B \cup C) = P(A) + P(A^c \cap B) + P(A^c \cap B^c \cap C)$$

and, in fact, for any finite number of events

$$P(A_1 \cup A_2 \cup A_3 \cup \ldots \cup A_n) = P(A_1) + P(A_1^c \cap A_2) + P(A_1^c \cap A_2^c \cap A_3) + \ldots + P(A_1^c \cap A_2^c \cap \cdots \cap A_{n-1}^c \cap A_n)$$

Remark.

$$\begin{aligned} P(A \cup B \cup C) &= P(C) + P(C^c \cap B) + P(C^c \cap B^c \cap A) \\ &= P(B) + P(B^c \cap C) + P(B^c \cap C^c \cap A) \\ &= \ldots \end{aligned}$$

<u>Conditional probability</u> is a probability where we are given partial information about the outcome of an experiment.

<u>Example</u> Roll two <u>balanced</u> 4-sided dice.

(1,1)	(1,2)	(1,3)	(1,4)
(2,1)	(2,2)	(2,3)	(2,4)
(3,1)	(3,2)	(3,3)	(3,4)
(4,1)	(4,2)	(4,3)	(4,4)

F

Ω

Sample space of this experiment represented in a Venn diagram

Dice are balanced $\Rightarrow$ 16 outcomes are equally-likely.

Let F be the event that the sum of the upfaces is 5.

Let G be the event that a 4 is rolled (at least once).

$P(F) = \frac{4}{16} = \frac{1}{4}$. $\quad P(G) = \frac{7}{16}$.

Suppose when the experiment is performed we observed the event G and now want to know the probability of F.

$P(F|G) = \frac{2}{7}$

pronounced conditional prob. of F Given G.

Once we know G occurred the sample space is reduced from Ω to G and G now has 7 equally-likely outcomes only 2 of which sum to 5.

In this last example $P(F|G) = \frac{2}{7}$ can also have been calculated as

$$P(F|G) = \frac{P(F\cap G)}{P(G)} = \frac{\frac{2}{16}}{\frac{7}{16}}$$ (we assume here that $P(G) > 0$.)

In general, if we have an experiment and we observe an event B with $P(B) > 0$, then we can define the (conditional) probability of A given B as

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

For fixed B, $P(\cdot|B)$ is a probability law:
Here's a check:

Axiom 1. Let $A \subseteq \Omega$, then $P(A|B) = \frac{P(A\cap B)}{P(B)} \geq 0$

since $P(\cdot)$ is nonnegative, and $P(B) > 0$.

Axiom 2: Let $A_1, A_2 \subseteq \Omega$ be disjoint. Then

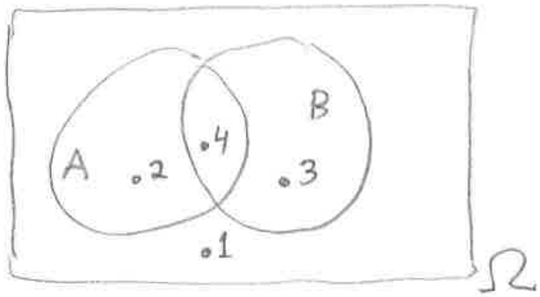
$$P(A_1 \cup A_2 | B) = \frac{P([A_1\cup A_2]\cap B)}{P(B)} = \frac{P((A_1\cap B)\cup(A_2\cap B))}{P(B)}$$

$$= \frac{P(A_1\cap B)}{P(B)} + \frac{P(A_2\cap B)}{P(B)} = P(A_1|B) + P(A_2|B)$$

Axiom 3

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

Example Consider the events $A, B \subseteq \Omega$ pictured below:



We are given the probabilities in each disjoint subregion

Compute each of the following:

$$P(A \mid B) = \frac{.4}{.4+.3} = \frac{4}{7}$$

$$P(B \mid A) = \frac{.4}{.2+.4} = \frac{2}{3}$$

$$P(A \mid B^c) = \frac{.2}{.2+.1} = \frac{2}{3}$$

$$P(B \mid A^c) = \frac{.3}{.1+.3} = \frac{3}{4}$$

$$P(A \mid A \cup B) = \frac{.6}{.2+.4+.3} = \frac{2}{3}$$

$$P(A \mid A \cup B^c) = \frac{.6}{.2+.4+.1} = \frac{6}{7}$$

∎

Example A box has 2 blue, and 3 green marbles — Call them b_1, b_2, g_1, g_2, g_3 and experiment is to draw 2 marbles at once.

Given that you've drawn a green marble, what is the conditional probability they are both green?

Define G to be the event that at least one of the two marbles is green.

D the event that both are green.

$$\Omega = \{ \{b_1,b_2\}, \{b_1,g_1\}, \{b_1,g_2\}, \{b_1,g_3\}, \{b_2,g_1\}, \{b_2,g_2\}, \{b_2,g_3\}, \{g_1,g_2\}, \{g_1,g_3\}, \{g_2,g_3\} \}$$

are all equally-likely.

$$P(G) = \frac{9}{10}, \quad P(D \cap G) = \frac{3}{10} \Rightarrow P(D|G) = \frac{\frac{3}{10}}{\frac{9}{10}} = \frac{1}{3}.$$

In the last two examples either the probability space is specified or (unconditional) probabilities we specified and we used

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

to compute conditional probabilities.

In many situations we proceed in the opposite direction. That is, we are given in advance some conditional probabilities (or what we want some conditional probabilities to be) and use this information to compute unconditional probabilities and possibly other conditional probabilities.

Here is a typical example of this situation

[We will use $P(A|B) = P(B)P(A|B)$ for events A, B with $P(B) > 0$]

Suppose that the population of a certain city is 60% female and 40% male. Suppose also that 30% of females and 50% of males smoke.

(a) What percentage are female smokers?

(b) What percentage of smokers are female?

(c) What percentage of nonsmokers are male?

To fix notation, let's set

F to be the set of females (F^c set of males)

S to be the set of smokers (S^c set of nonsmokers)

With this notation:

in (a) we are looking for $P(F \cap S)$

in (b) we are looking for $P(F \mid S)$

in (c) we are looking for $P(F^c \mid S^c)$

From the problem statement we are given:

$$P(F) = .60, \quad P(F^c) = .40, \quad P(S \mid F) = .30, \quad P(S \mid F^c) = .50$$

(a) By the condition probability formula we have

$$P(F \cap S) = P(F) \cdot P(S \mid F) = .60(.30) = .18.$$

It is also true that $P(F \cap S) = P(S)P(F \mid S)$ but the two probabitities on the right were not given to us directly

(b) $P(F \mid S) = \dfrac{P(F \cap S)}{P(S)} = \dfrac{.18}{P(S)}$ and we now need to compute $P(S)$.

are disjoint

$$S = (S \cap F) \cup (S \cap F^c)$$

So $P(S) = P(S \cap F) + P(S \cap F^c)$

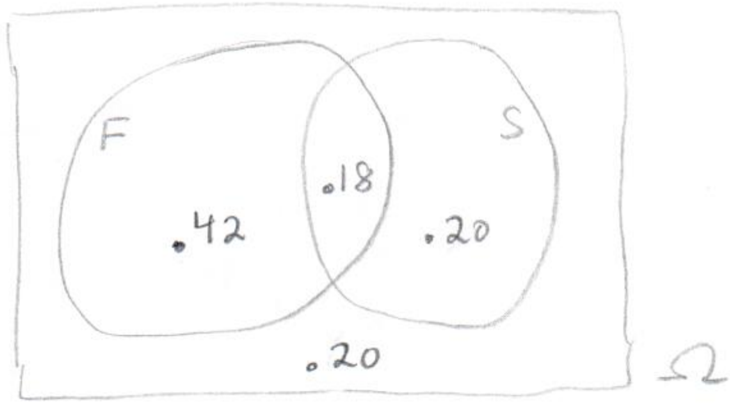
$$= P(F)P(S \mid F) + P(F^c)P(S \mid F^c) = \underbrace{.60(.30) + .40(.50)}_{= .38}$$

Therefore, $P(F \mid S) = \dfrac{.18}{.38} = \dfrac{9}{19}$.

(c) $P(F^c|S^c) = \dfrac{P(F^c \cap S^c)}{P(S^c)} = \dfrac{P(F^c)P(S^c|F^c)}{1-P(S)} = \dfrac{P(F^c)(1-P(S|F^c))}{1-P(S)}$

$$= \frac{(.4)(1-.50)}{1-.38} = \frac{.20}{.62} = \frac{10}{31}.$$

A Venn diagram of the situation could also have been constructed at the beginning of the solution stage:



We put the probability of .18 into the intersection of F and S from part (a). i.e. since $P(S|F) = \frac{P(S \cap F)}{P(F)} = .3 \Rightarrow P(S \cap F) = .3\,P(F) = .18$.
Now, since $P(F) = .6$ it must be that $.60 - .18 = .42$ probability of $F \cap S^c$.

Next, since $P(S|F^c) = \dfrac{P(S \cap F^c)}{\underbrace{P(F^c)}_{.4}} = .5 \Rightarrow P(S \cap F^c) = .5(.4) = .2$

Lastly, $P(\Omega) = 1$ so

$$P(F^c \cap S^c) = 1 - (.42 + .18 + .20) = .20.$$

and the probabilities in parts (a), (b) and (c) can hence be computed.

Multiplicative rule(s) of conditional probability

$$P(A_1 \cap A_2) = P(A_1)\, P(A_2 | A_1)$$

$$(*) \quad P(A_1 \cap A_2 \cap A_3) = P(A_1)\, P(A_2 | A_1)\, P(A_3 | A_1 \cap A_2)$$

$$P(A_1 \cap A_2 \cap A_3 \cap A_4) = P(A_1)\, P(A_2 | A_1)\, P(A_3 | A_1 \cap A_2)\, P(A_4 | A_1 \cap A_2 \cap A_3)$$

In general,

$$P\left(\bigcap_{i=1}^{n} A_i\right) = P(A_1) \cdot P(A_2 | A_1) \cdot P(A_3 | A_1 \cap A_2) \cdots P\left(A_n \Big| \bigcap_{i=1}^{n-1} A_i\right)$$

To see why this is true, we'll show $(*)$ above:

$$P(A_1)\, P(A_2 | A_1)\, P(A_3 | A_1 \cap A_2)$$

$$= \cancel{P(A_1)} \cdot \frac{\cancel{P(A_1 \cap A_2)}}{\cancel{P(A_1)}} \cdot \frac{P(A_1 \cap A_2 \cap A_3)}{\cancel{P(A_1 \cap A_2)}} = P(A_1 \cap A_2 \cap A_3) \quad \checkmark$$

In the previous example we saw an illustration of the multiplicative rule with 2 sets.

Let's now do an example using the multiplicative rule with 3 sets...

<u>Example</u> A box contains 10 marbles comprised of 4 blue and 6 green marbles.

Consider the experiment where 3 marbles are drawn one-at-a time <u>without replacement</u> (and we observe the color on each draw.)

(a) What's the probability of drawing 3 (consecutive) blue marbles?

(b) What's the probability of drawing 3 (consecutive) green marbles?

(c) What's the probability of exactly one blue drawn?

<u>Solution</u>: Let B_i = blue marble pulled on i^{th} draw

G_i = green " "

(a) $P(B_1 \cap B_2 \cap B_3) = P(B_1)\,P(B_2 \mid B_1)\,P(B_3 \mid B_1 \cap B_2)$

$$= \frac{4}{10} \cdot \frac{3}{9} \cdot \frac{2}{8} = \frac{24}{720}$$

(b) $P(G_1 \cap G_2 \cap G_3) = P(G_1) \cdot P(G_2 \mid G_1) \cdot P(G_3 \mid G_1 \cap G_2)$

$$= \frac{6}{10} \cdot \frac{5}{9} \cdot \frac{4}{8} = \frac{120}{720}.$$

5 times more likely that 3 blues.

(c) The event 'exactly one blue marble' doesn't say on which draw we observe it, that is, each of the events

$$B_1 \cap G_2 \cap G_3 \quad , \ G_1 \cap B_2 \cap G_3 \ \text{ and } \ G_1 \cap G_2 \cap B_3$$

results in exactly one blue marble (and these are the only ways exactly one blue marble can be drawn).

Therefore, the event 'exactly one blue marble' is the disjoint union

$$(B_1 \cap G_2 \cap G_3) \cup (G_1 \cap B_2 \cap G_3) \cup (G_1 \cap G_2 \cap B_3)$$

Moreover,

$$P(B_1 \cap G_2 \cap G_3) = P(B_1) P(G_2 | B_1) \ P(G_3 | B_1 \cap G_2)$$
$$= \frac{4}{10} \cdot \frac{6}{9} \cdot \frac{5}{8} = \frac{120}{720}$$

$$P(G_1 \cap B_2 \cap G_3) = P(G_1) \ P(B_2 | G_1) \cdot P(G_3 | G_1 \cap B_2)$$
$$= \frac{6}{10} \cdot \frac{4}{9} \cdot \frac{5}{8} = \frac{120}{720}$$

Similarly, $P(G_1 \cap G_2 \cap B_3) = \frac{120}{720}$ Thus the probability of exactly one blue marble is

$$\frac{120}{720} + \frac{120}{720} + \frac{120}{720} = \frac{360}{720} = \frac{1}{2}.$$

Remark

In the last example, it's fairly plain that $P(B_1) = \frac{4}{10} = \frac{2}{5}$.

What's $P(B_2)$?

To answer this question we write

$$B_2 = (B_2 \cap B_1) \cup (B_2 \cap B_1^c)$$
$$= (B_2 \cap B_1) \cup (B_2 \cap G_1).$$

are disjoint

$$P(B_2) = P(B_2 \cap B_1) + P(B_2 \cap G_1)$$
$$= P(B_1)P(B_2|B_1) + P(G_1)P(B_2|G_1)$$
$$= \frac{4}{10} \cdot \frac{3}{9} + \frac{6}{10} \cdot \frac{4}{9} = \frac{36}{90} = \frac{4}{10} = \frac{2}{5}$$

So, $P(B_2) = P(B_1)$!!!

Therefore, $P(G_2) = 1 - P(B_2) = \frac{6}{10} = P(G_1)$ as well.

How about $P(B_3)$?

$$B_3 = (B_3 \cap B_2) \cup (B_3 \cap G_2)$$

$$P(B_3) = P(B_2)P(B_3|B_2) + P(G_2)P(B_3|G_2)$$
$$= \frac{4}{10} \cdot \frac{3}{9} + \frac{6}{10} \cdot \frac{4}{9} = \frac{2}{5}.$$

$\boxtimes$