A. **Axioms of Probability.** Recall, a **sample space** Ω is a set of outcomes.

For example, consider a weighted 6-sided die, with even numbers twice as likely to appear as odd numbers. The set of outcomes of a single roll is:

sample space $\implies \Omega = \{1, 2, 3, 4, 5, 6\}$

Subsets of Ω are called **events**, and are assigned **nonnegative** probabilities:

$\mathbb{P}(\text{the roll has a value from 1 to 6}) = \mathbb{P}(\{1, 2, 3, 4, 5, 6\}) = \mathbb{P}(\Omega) =$

$\mathbb{P}(\text{the roll has an odd value}) = \mathbb{P}(\{1, 3, 5\}) =$

$\mathbb{P}(\text{the roll has an even value}) = \mathbb{P}(\{2, 4, 6\}) =$

The events of even and odd values are said to be **mutually exclusive** because they are **disjoint**. Further:

Recall, disjoint meant their intersection was empty.

$\Omega =$

> **Kolmogorov Axioms of Probability.**
>
> A **probability space** consists of a sample space Ω, an **event space** made up of particular subsets of Ω called events, and a probability function $\mathbb{P}$ that assigns probabilities to events, such that the following axioms are satisfied.
>
> **Nonnegativity:** For any event E, $\mathbb{P}(E) \geqslant 0$.
>
> **Total Probability:** $\mathbb{P}(\Omega) = 1$.
>
> **Countable Additivity:** For countably many events $E_1, E_2, \ldots$ that are **mutually exclusive**, meaning any pair of them is **disjoint**, then:
>
> $$\mathbb{P}\left(\bigsqcup_i E_i\right) = \sum_i \mathbb{P}(E_i).$$

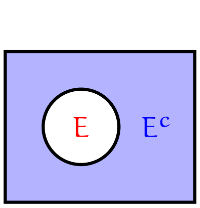
When the sample space is countable, the event space consists of **all** subsets of the sample space. If the sample space is not countable, the situation is more subtle.

This includes countably infinite, or finitely many.

Remember that $\bigsqcup$ is an indicator that the union is of disjoint sets. More precisely, when used this manner, the sets should be pairwise disjoint, meaning any pair of them is disjoint.

The **complement** of an event E in the sample space Ω is defined to be:

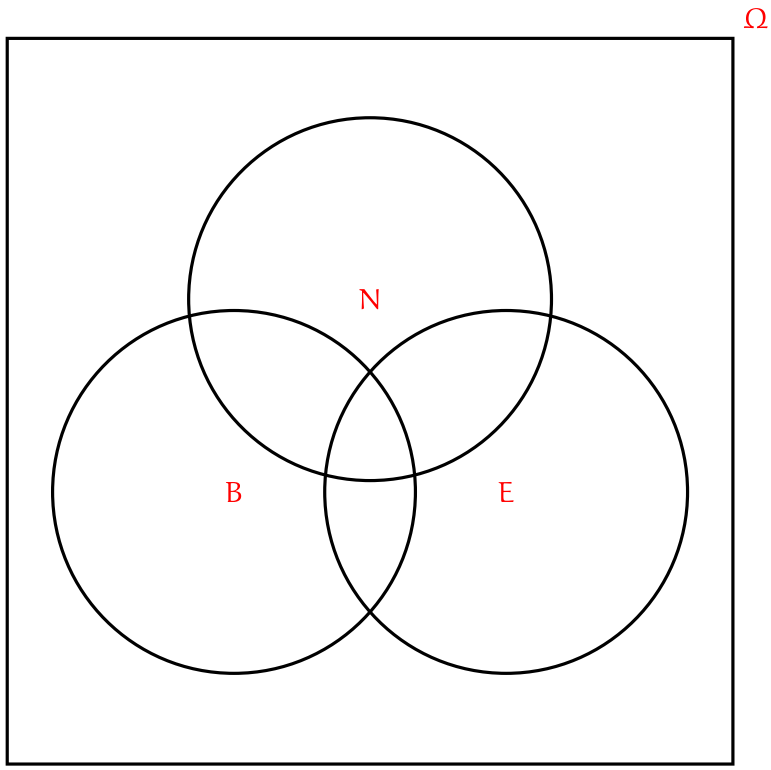
$E^c = \Omega \setminus E = \{x \in \Omega \mid x \notin E\}$



> **Complementary Probability.** For any event E:
>
> $\mathbb{P}(E^c) =$

Example 1. At a certain school, students wear accessories from among: necklaces, earrings, and bracelets. 10% of students wear all three, 20% wear a necklace and a bracelet, 20% wear a necklace and earrings, 15% wear an earrings and a bracelet, and 87% wear at least one of these accessories. The percent of students that wears bracelets is equal to the percent that wears earrings, but is $1/2$ the percent that wears necklaces.

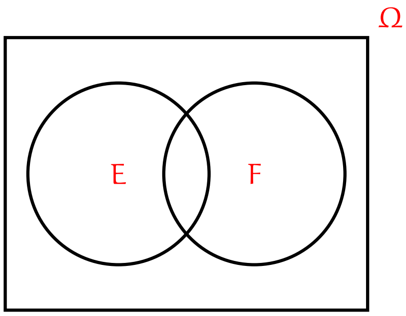
If a student is selected uniformly at random, find the probability of each of the events N that the student wears a necklace, E that the student wears earrings, and B that the student wears a bracelet.



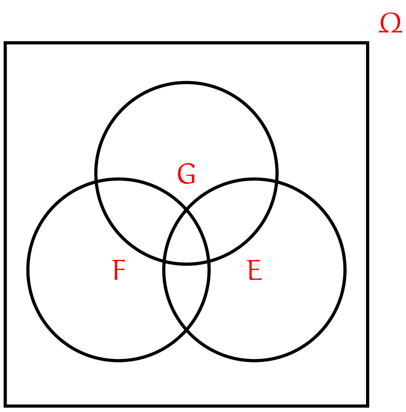
B. **Inclusion–Exclusion Principle.** The law of countable additivity tells us the probability of a union of **mutually exclusive** events. What if they are not mutually exclusive?

$\mathbb{P}(E \cup F) =$

When you merely add the probabilities of E and F, you count the probability of the intersection EF twice!



$\mathbb{P}(E \cup F \cup G) =$



When you add the probabilities of E and F and G and substract by the probabilities of the interesctions EF, EG, FG, you have first overcounted the probability of the triple intersection three times, and then subtracted the probability of the triple intersection three times! You still need to account for the probability of the triple intersection!

This process can be extended for any finite number of events.

Inclusion–Exclusion. The probability of a union of events $E_1, \ldots, E_n$ is:

$$\begin{aligned}
\mathbb{P}\left(\bigcup_{i=1}^{n} E_i\right) = &+ \sum_{i=1}^{n} \mathbb{P}(E_i) \\
&- \sum_{1 \leqslant i_1 < i_2 \leqslant n} \mathbb{P}(E_{i_1} E_{i_1}) \\
&+ \sum_{1 \leqslant i_1 < i_2 < i_3 \leqslant n} \mathbb{P}(E_{i_1} E_{i_2} E_{i_3}) \\
&\vdots \\
&+ (-1)^{n+1} \mathbb{P}(E_1 E_2 \ldots E_n)
\end{aligned}$$

or more compactly:

$$\mathbb{P}\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{r=1}^{n} \left[(-1)^{r+1} \sum_{1 \leqslant i_1 < i_2 < \cdots < i_r \leqslant n} \mathbb{P}(E_{i_1} E_{i_2} \cdots E_{i_r}) \right]$$

While this looks complex, it simply says, to compute the probability of the union, add the probabilities of each event, subtract the probabilities of all double intersections, add the probabilities of all triple intersections, and so on.

It can help with calculations to note that the number of terms in each sum:

$$\sum_{1 \leqslant i_1 < i_2 < \cdots < i_r \leqslant n}$$

is the number of ways of choosing r numbers from n, or in other words is $\binom{n}{r}$.

Example 2. Solve this problem again, using inclusion–exclusion.

At a certain school, students wear accessories from among: necklaces, earrings, and bracelets. 10% of students wear all three, 20% wear a necklace and a bracelet, 20% wear a necklace and earrings, 15% wear an earrings and a bracelet, and 87% wear at least one of these accessories. The percent of students that wears bracelets is equal to the percent that wears earrings, but is $1/2$ the percent that wears necklaces.

If a student is selected uniformly at random, find the probability of each of the events N that the student wears a necklace, E that the student wears earrings, and B that the student wears a bracelet.

C. **Equally Likely Outcomes.** Recall, on the first day of class we talked about sample spaces with equally likely outcomes, like the outcomes of a roll of a **fair** die. Here is how you calculate probability for such sample spaces.

Equally Likely Outcomes. A **finite** sample space Ω with **equally likely** outcomes is said to have a **discrete uniform** probability distribution, meaning that the probablity of an event $A \subseteq \Omega$ is given by:

$\mathbb{P}(A) =$

The implication here is that this genuinely defines a probability space, meaning all the axioms are satisfied.

For a **discrete uniform** distribution, the probability of a single outcome $x \in \Omega$ is:

$\mathbb{P}(\{x\}) =$

So, for example, in the roll of fair 6–sided die, the probability of any value is $\frac{1}{6}$.

Example 3. Two fair 6–sided dice are rolled. What is the probability that the sum is 8?

Be careful, if you select as your sample space the set of values of the sums, then the outcomes are not equally likely. This is why, here, it is better to use the set of equally likely outcomes for ordered pairs of values on the rolls.

Note, we cannot just use stars and bars to count the number of positive integer solutions to $x_1 + x_2 = 8$. Why? For example, $(7, 1)$ is a solution, but is not a valid pair of values from dice.