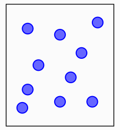
Probability

Probability Models and Axioms

Sample Space

Let the sample space $\Omega = \{\omega_1, \ldots, \omega_n\}$ be the set of all possible outcomes of an experiment or random trial. Let $\omega_i \in \Omega$ be a particular sample point or outcome.

In order to be a sample space, the set $\{\omega_1, \ldots, \omega_n\}$ must meet certain conditions:



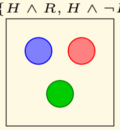
Outcomes must be **Mutually Exclusive**, i.e. if ω_i occurs, then no other ω_j will take place. $\forall i, j = 1 \ldots n, i \neq j$.

Outcomes must be **Collectively Exhaustive**, i.e. on every experiment or trial, the outcome must be only $\omega_i \in \Omega$.

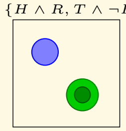
The set must be at a **Right Granularity**, depending on what the experimenter is interested in. Irrelevant information must be removed from the set and the right abstraction must be chosen.

Consider the experiment of flipping a coin. Let H and T be the events heads and tails, respectively. Let R be "it's raining" and $\neg R$ its negation.

$\Omega = \{H \wedge R, H \wedge \neg R, T\}$



$\Omega = \{H \wedge R, T \wedge \neg R, T\}$



For the left case, we have a legitimate sample space. For the right case we have an outcome that it's included in another one, namely $\{T \wedge \neg R\} \subset T$, therefore the outcomes aren't mutually exclusive.

Discrete and Continuous, Finite and Infinite Sample Spaces

- Discrete and Finite → Throwing 2 regular dice once: $|\Omega| = 6 \times 6 = 36$
- Discrete and Infinite → Guessing a natural number (?)
- Continuous and Infinite → The (x,y) coordinates of the landing of a dart

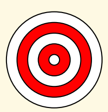
Events

An event A is a **subset** of the sample space Ω, $\rightarrow A \subset \Omega$, i.e. an event A is a set of outcomes itself. Probability is assigned to events, $P(A)$. If events weren't defined in terms of subsets (sets), handling individual sample points in continuous sample spaces would be complicated.

Probability Axioms

Nonnegativity: $P(A) \geq 0$, A is any event.
Normalization: $P(\Omega) = 1$ (the event here is Ω, since every set is subset of itself).
(Finite) Additivity (*to be strengthened later*): $A \cap B = \emptyset \implies P(A \cup B) = P(A) + P(B)$

The last axiom is going to be refined later
These 3 axioms are sufficient to have a legitimate **probability model**.



Suppose the distribution of dart hits is uniform across the dart board, so the probability of hitting a point is $P(x, y) = \frac{1}{\text{Area}}$ over the region, and 0 elsewhere. Since a single point has no area, we can express it as a region approaching zero, $R \to 0$. Therefore, the probability of **exactly hitting the center** of an infinite continuous space is:

$$P\left[(x, y) = (x_0, y_0)\right] = \lim_{R \to 0} \iint_R f(x, y)\, dx\, dy = 0$$

since the integral over an infinitesimally small region R will go to zero as R goes to zero. But if we consider the event of hitting a region instead of a point, its probability is greater than zero.

Probability Model

Bertsekas & Tsitsiklis definition: A Probabilistic Model is a mathematical description of an uncertain situation. It is composed of two main *ingredients*: A Sample Space and a Probility Law that specifies the likelihood of events.

Probability Law: The logic by which likelihood of outcomes is defined or assigned.
Consider the probability of hitting any subset of a 1×1 square. $P(x, y) : 0 \leq x, y \leq 1 \rightarrow$ The probability of any particular subset of Ω is just its area (Uniform Probability).

For completeness (Sample Space, **Probability Space**, Probability Model):
-https://stats.stackexchange.com/questions/199280/why-do-we-need-sigma-algebras-to-define-probability-spaces
-https://math.stackexchange.com/questions/2002416/defining-the-sigma-algebra-of-events-of-a-probability-space
-https://en.wikipedia.org/wiki/Probability_space

Countable and Uncountable Sets

Segun Tsitsiklis, Discrete = Countable, alrededor del min 10:24, en el video 17, Lec. 1.

Consequences of the Axioms

By set theory definitions we have: $\boxed{A \cup A^c = \Omega}$ and $\boxed{A \cap A^c = \emptyset}$

$$\boxed{P(A) \leq 1}$$

A and A^c are disjoint $\Rightarrow P(A \cup A^c) = 1 = P(A) + P(A^c) \Rightarrow P(A^c) = 1 - P(A)$, and by *nonnegativity* we get $P(A^c) \geq 0 \Rightarrow P(A) \leq 1$ ■

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

Let $A = \Omega \Rightarrow P(\Omega) + P(\Omega^c) = 1 \Rightarrow 1 + \emptyset = 1 \Rightarrow P(\emptyset) = 0$ ■

Let Ω be a finite set and $A_1, \ldots, A_n$ be disjoint events, then:

$$\boxed{P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{1}^{n} P(A_i)}$$

$P(A \cup B \cup C) = P[(A \cup B) \cup C]$. From additivity, given that the events are disjoint, we have $(P(A) + P(B)) + P(C)$. By induction we can extend this to n disjoint sets ■

Let $\{\omega_1, \ldots, \omega_k\}$ be a discrete, finite set of sample points, then:

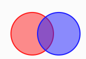
$$\boxed{P\left(\{\omega_1, \ldots, \omega_k\}\right) \Rightarrow P\left(\bigcup_{j=1}^{k} \{\omega_j\}\right) \Rightarrow \sum_{j=1}^{k} P\left(\{\omega_j\}\right)}$$

because $\{\omega_1, \ldots, \omega_k\}$, can be seen as the union of *unit sets*, and since they are disjoint, additivity applies ■. Although, a simpler, non rigorous notation can be used: $\sum_{j=1}^{k} P(\omega_j)$.

Let A, B, C be disjoint subsets of Ω, then:
$P(A) + P(A^c) + P(B) \neq P(A \cup A^c \cup B) \rightarrow$ e.g. when: $A = \emptyset$, $B = \Omega$
$P(A^c) + P(B) \nleq 1 \rightarrow$ e.g. when: $A = \emptyset$, $B = \Omega \Rightarrow A^c = \Omega$
Let A, B, C be not necessarily disjoint subsets of Ω, then:
$P[(A \cap B) \cup (C \cap A^c)] \leq P(A \cup B \cup C) \rightarrow$ Since $(A \cap B)$, $(C \cup A^c)$ are its subsets.

More Consequences of the Axioms

Consider the condition $P(A \cap B) \geq 0$, $\Rightarrow$ The events could be joint, therefore, more generally:



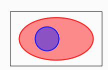
$$\boxed{P(A \cup B) = P(A) + P(B) - P(A \cap B)}$$

Which can be generalized to the Inclusion-Exclusion Principle:

$$\boxed{P\left(\bigcup_{i=1}^{n} A_i\right) = -\sum_{k=1}^{n} (-1)^k \sum_{1 \leq i_1 < \ldots < i_k \leq n} P\left(\bigcap_{j=1}^{k} A_{i_j}\right)}$$

From the above, the *Union Bound* property follows: $\boxed{P(A \cup B) \leq P(A) + P(B)}$

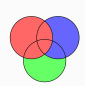
Consider that A is included in B, then:



$$\boxed{A \subset B \Rightarrow P(A) \leq P(B)}$$

since $B = A \cup (B \cap A^c) \Rightarrow P(B) = P(A) + P(B \cap A^c) \geq P(A)$ ■

Consider 3 sets not necessarily disjoint, e.g.:



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

Visually, we can check the boxed expression by the matching of the colors, and since the subsets are disjoint, additivity holds. Notice the expression also applies to disjoint sets ■

Pendiente...

- Discrete Probability Law & Discrete Uniform Probability Law → Textbook. A■adir adem■s un ejemplo similar al de la secci■n 14
- Continuous Probability Low → https://stats.stackexchange.com/questions/273382/how-can-the-probability-of-each-point-be-zero-in-continuous-random-variable. Incluir ademas algun ejemplo algo complejo como los de la seccion 16.

Countable Additivity Axiom

If $A_1, A_2, \ldots$ is an infinite **sequence** of disjoint events, then:

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

The word sequence is important here since it's necessary to be able to arrange the events in some order.

Consider the following case: The sample space consists of the unit square, the probability of a set/event is its area. Now, consider the probability of the whole Ω as the probability of the union of all the (x, y) points: $P(\Omega) = 1 = P\left(\bigcup \{(x, y)\}\right) = \sum P(\{x, y\}) = \sum 0 = 0$, we arrived to a seemingly contradiction because the elements of the unit square (i.e. (x, y) sets) can't be arranged in a sequence → The unit square is an uncountable set.

The proof for "countable/discrete = can be arranged in a sequence, uncountable/continous = can't be arranged in a sequence" is said to be found in Measure Theory.

Ejemplo de sumatoria de serie infinita (ver ejercicios 18, 19, 20)

Let the sample space be the set of all positive integers. Is it possible to have a "uniform" probability low, that is, a probability law that assigns the same probability c to each positive integer?

Suppose that $c = 0$. Then: $1 = \mathbf{P}(\Omega) = \mathbf{P}(\{1\} \cup \{2\} \cup \{3\} \ldots)$, and by countable additivity this equals $\mathbf{P}(\{1\}) + \mathbf{P}(\{2\}) + \mathbf{P}(\{3\}) + \ldots = \sum 0 = 0$, which is a contradiction.

Suppose that $c > 0$. Then, there exists an integer k such that $kc > 1$. By additivity, $P(1, 2, \ldots, k) = kc > 1$. The answer is therefore **No**.

Important: The question ask whether is possible to have a "uniform" probability law (each event has the same probability) from this discrete, countable set, not whether **countable additivity** can be applied, which implies that a non-uniform probability law could be still applied.

Probability and Statistics Relationship

Insert the last figure from the PDF

Conditioning and Independence

Conditioning and Bayes' Rule

Conditional Probability

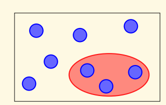
Let $P(B) > 0$, then the probability of A given that B has occurred is:

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

Conditional probabilities share properties of ordinary probabilities:
$P(A|B) \geq 0$, assuming $P(B) > 0$
$P(\Omega|B) = \frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1 \rightarrow P(B|B)$ has the same result.
If $A \cap C = \emptyset \Rightarrow P(A \cup C|B) = P(A|B) + P(C|B)$, because
$P(A \cup C|B) = \frac{P[(A \cup C) \cap B]}{P(B)} = \frac{P[(A \cap B) \cup (C \cap B)]}{P(B)} = \frac{P(A \cap B) + P(C \cap B)}{P(B)}$,
and by induction this can be proven true for finitely many disjoint events (**Finite Additivity**) and countably many disjoint events (**Countable Additivity**).

Any fact we derive for ordinary probability will remain true for conditional probability as well.

Consider a finite Ω with a discrete uniform probability law. Let $B \neq \emptyset$, e.g.:



The conditional probability law on B, given that B ocurred, is also discrete uniform. Each event inside B would have a probability of $\frac{1}{|B|}$.

The conditional probability law on Ω, given that B ocurred, is not discrete uniform. Events outside B have 0 probability, different from events inside.

Multiplication Rule

Notice that:

$$\begin{aligned} P(A \cap B) &= P(B)P(A|B) \\ &= P(A)P(B|A) \end{aligned}$$

And for 3 events we have:

$$\begin{aligned} P[(A \cap B) \cap C] &= P(A \cap B)P(C|A \cap B) \\ &= P(A)P(B|A)P(C|A \cap B) \end{aligned}$$

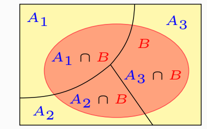
More generally:

$$\boxed{P\left(\bigcap_{i=1}^{n} A_i\right) = P(A_1) \prod_{i=2}^{n} P\left(A_i \,\middle|\, \bigcap_{j=1}^{i-1} A_j\right)}$$

A particular intersection of events would be represented as a full path in a probability tree.

Total Probability Rule

- Consider a partition of Ω into A_i events. Since it's a partition, events are disjoint.
- Let's say we have a probability model for A_i.
- We have also modeled $P(B)$ for each scenario, i.e. $P(B|A_i)$
- We can use **finite additivity** in order to calculate $P(B)$



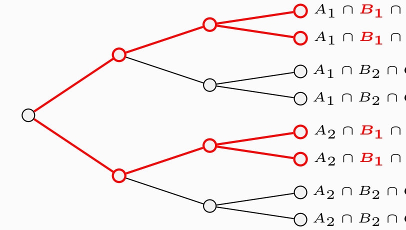
$$\begin{aligned} P(B) &= P(B \cap A_1) + P(B \cap A_2) + P(B \cap A_3) \\ &= P(A_1)P(B|A_1) + \ldots + P(A_3)P(B|A_3) \end{aligned}$$

By induction, for a *2-level probability scenario*, it can be proven that:

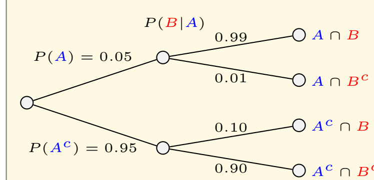
$$\boxed{P(B) = \sum_{i=1} P(A_i)P(B|A_i)}$$

Multiplication & Total Probability Rules: Tree Interpretation

Consider the following 3-level probability tree (scenario): A_1, A_2 are disjoint, B_1, B_2 are disjoint, C_1, C_2 are disjoint, then B_1's total probability is calculated using both rules:



Consider the following events: A : Airplane is flying above, B : Something registers on the radar screen. Some conditional probabilities are depicted in the figure, e.g. $P(B|A) = 0.99$. Find the probability that there's an airplane flying above, given that the radar registers something.



$$\begin{aligned} P(A \cap B) &= P(A)\,P(B|A) \\ &= (0.05)(0.99) \end{aligned}$$

$$\begin{aligned} P(B) &= (0.05)(0.99) + \\ &\quad (0.95)(0.10) \\ &= 0.1445 \end{aligned}$$

$$\begin{aligned} P(A|B) &= \frac{(0.5)(0.99)}{0.1445} \\ &= 0.34 \end{aligned}$$

Bayes' Rule

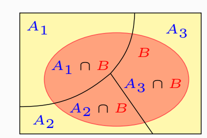
- Consider a partition of Ω into A_i disjoint events.
- We have an initial model or beliefs for A_i.
- We know how likely it is a particular event B under each scenario, i.e. $P(B|A_i)$.

$$A_i \xrightarrow[P(B|A_i)]{\text{model}} B$$

- Given that B occurred, we can update our model: Analyze possible causes or most likely scenarios for B.

$$B \xrightarrow[P(A_i|B)]{\text{inference}} A_i$$

- In other words, we use inference to analyze how likely is a scenario A_i, given that B occurred.



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

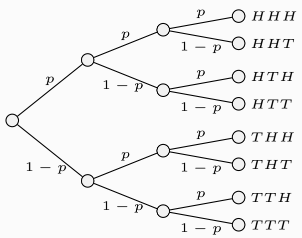
$$\boxed{P(A_i|B) = \frac{P(A)P(B|A_i)}{\sum_j P(A_j)P(B|A_j)}}$$

Check the diagrams: https://en.wikipedia.org/wiki/Bayes%27_theorem#Random_variables

Independence

When Conditional Probability = Unconditional Probability

Consider 3 tosses of a biased coin: $P(H) = p$, $P(T) = 1 - p$



Notice that the unconditional and conditional probabilities for H_2 (heads in the second toss) are the same:

Conditional Probability (from the diagram)

$$P(H_2|H_1) = P(H_2|T_1) = p$$

Unconditional Probability (using Total Probability)

$$P(H_2) = P(H_1)P(H_2|H_1) + P(T_1)P(H_2|T_1) = (p)(p) + (1-p)(p) = p$$

Independence of Two Events

Intuitive definition: $P(B|A) = P(B) \Rightarrow$ The ocurrence of A gives no new information about B.

Formal definition:

Two events are independent if:

$$\boxed{P(A \cap B) = P(A)P(B)}$$

This definition symmetrically implies both $P(B|A) = P(B)$ and $P(A|B) = P(A)$ and works even when $P(A) = 0$.

Why $P(A \cap B) = P(A)P(B) \iff$ Independence

So far we know the following:

$$\text{Independent}(A, B) \Rightarrow P(A \cap B) = P(A|B)P(B) = P(A)P(B)$$
$$\text{Dependent}(A, B) \Rightarrow P(A \cap B) = P(A|B)P(B) \neq P(A)P(B)$$

If by pure coincidence we have

$$\text{Dependent}(A, B) \wedge P(A \cap B) = P(A)P(B)$$

Then that would imply both:

$$P(A|B)P(B) = P(A)P(B)$$
$$P(A|B)P(B) \neq P(A)P(B) \quad \text{(a contradiction)}$$

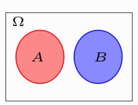
Independence of "Essentially Deterministic" Events

$\boxed{P(A) = 0 \vee P(A) = 1 \Rightarrow A \text{ is independent of any event and independent of itself.}}$

Let B be any event $\Rightarrow P(A \cap B) = P(A)P(B)$ for both cases $P(A) = 0 \wedge P(A) = 1$ ■

Independence and Disjoint Sets

- Let A and B be disjoint events $\Rightarrow P(A \cap B) = 0$.
- Let $P(A), P(B) > 0 \Rightarrow P(A)P(B) > 0$
- Therefore $P(A \cap B) = P(A)P(B)$ does not hold $\Rightarrow$ Events are not independent.



Being independent is something completely different from being disjoint. Furthermore, for mutually exclusive, disjoint events with positive probabilities, if A occurs $\Rightarrow$ B can't occur, so it gives us a lot of information. Therefore:

Disjoint events are not independent

Independence is a relation about information

If one of the events is impossible $\Rightarrow$ They are **trivially independent**.

We have a peculiar coin. When tossed twice, the first toss results in H with probability $1/2$. However, the second toss always yields the same result as the first toss. Thus, the only possible outcomes for a sequence of 2 tosses are HH and TT, and both have equal probabilities. Are the two events $A = H_1$ and $B = H_2$ independent?

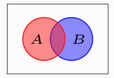
$P(A) = P(B) = P(A \cap B) = 1/2 \Rightarrow P(A \cap B) \neq P(A)P(B) \Rightarrow$ Not independent.

Let $A \subset \Omega \Rightarrow P(A \cap \Omega) = P(A) \Rightarrow P(A \cap \Omega) = P(A)P(\Omega) \Rightarrow$ Independent.

Let $0 < P(A) < 1 \Rightarrow P(A \cap A) = P(A) \neq P(A)P(A) \Rightarrow$ Not independent.

Independence of Complements

Let A and B be independent, then:



$$A = (A \cap B) \cup (A \cap B^c)$$
$$P(A) = P(A \cap B) + P(A \cap B^c) \quad \text{(since they're disjoint)}$$
$$P(A) = P(A)P(B) + P(A \cap B^c)$$
$$P(A \cap B^c) = P(A) - P(A)P(B) = P(A)(1 - P(B))$$

$$\boxed{P\left(A \cap B^c\right) = P\left(A\right)P\left(B^c\right)}$$

By the same logic: A and B independent $\Rightarrow$ A and B^c independent $\Rightarrow$ B^c and A independent $\Rightarrow$ B^c and A^c independent $\Rightarrow$ A^c and B^c independent.

$$\boxed{P\left(A^c \cap B^c\right) = P\left(A^c\right)P\left(B^c\right)}$$

Conditional Independence

Given an event C, events A and B are conditionally independent if:

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

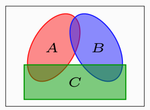
Now, notice that by the Bayes' Rule and Multiplication Rule:

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

So conditional independence also implies: $\boxed{P(A|B \cap C) = P(A|C)}$

In words, this relation states that if C is known to have occurred, the additional knowledge that B also occurred does not change the probability of A.

Independence does not imply Conditional Independence

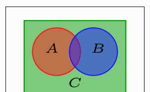


For any probabilistic model, let A and B be independent events, and let C be an event such that $P(C) > 0$, $P(A|C) > 0$, and $P(B|C) > 0$, while $A \cap B \cap C = \emptyset$

$\Rightarrow$ A and B can't be conditionally independent since $P(A \cap B|C) = 0$ while $P(A|C)P(B|C) > O$ ■.

Consider two independent fair coin tosses, in which all four possible outcomes are equally likely. Let H_1 = {1st toss is a head}, H_2 = {2nd toss is a head}, D = {the two tosses have different results}. Then events H_1 and H_2 are (unconditionally) independent, and $P(H_1|D) = P(H_2|D) = 1/2$, but since $P(H_1 \cap H_2|D) = 0 \Rightarrow P(H_1 \cap H_2|D) \neq P(H_1|D)P(H_2|D)$.

Conditional Independence does not imply Independence



Let A, B be conditionally independent given C:
$P(A \cap B|C) = P(A|C)P(B|C)$
$\Rightarrow P(A \cap B|C^c) = P(A|C^c)P(B|C^c)$.
By total probability we have:
$P(A) = P(C)P(A|C) + P(C^c)P(A|C^c)$ and
$P(B) = P(C)P(B|C) + P(C^c)P(B|C^c)$.

The unconditional probability of $A \cap B = P(C)P(A \cap B|C) + P(C^c)P(A \cap B|C^c)$, so $P(A \cap B) = P(C)P(A|C)P(B|C) + P(C^c)P(A|C^c)P(B|C^c)$, therefore $P(A \cap B) = P(A)P(B)$ is not necessarily true ■

Let A and B be conditionally independent given $C \Rightarrow P(A \cap B|C) = P(A|C)P(B|C)$.
Let $P(C) = 1/2$ and $P(A|C) = P(B|C) = 1/2$, and also $P(A|C^c) = P(B|C^c) = 0$.
So $P(A \cap B) = P(C)P(A \cap B|C) + P(C^c)P(A \cap B|C^c) = (1/2)^3 + 0 = 1/8$.

But
$P(A)P(B) = [P(C)P(A|C) + P(C^c)P(A|C^c)][P(C)P(B|C) + P(C^c)P(B|C^c)]$
$= (1/2)(1/2) = 1/4 \Rightarrow P(A \cap B) \neq P(A)P(B)$

Conditional Independence & Complements

Let A and B be **independent and conditionally independent** given C. Let $P(C), P(C^c) > 0$
$\Rightarrow$ A and B^c are guaranteed to be conditionally independent given C. Proof:

We need to get to the target equation: $P(A \cap B^c|C) = P(A|C)P(B^c|C)$. Notice that:
$P(B) + P(B^c) = 1 \Rightarrow P(B|C) + P(B^c|C) = 1 \Rightarrow P(B^c|C) = 1 - P(B|C)$.
Now, $(A \cap B)$ and $(A \cap B^c)$ are disjoint events (they can't both occur)
$\Rightarrow (A \cap B) \cup (A \cap B^c) = A$ (this can be proven using the properties of sets, but try to see it intuitively). Therefore $P(A \cap B^c|C) = P(A|C) - P(A \cap B|C)$.

Considering the assumptions, the target equation can now be expressed as:
$P(A|C) - P(A \cap B|C) = P(A|C)[1 - P(B|C)]$
$P(A|C) - P(A|B)P(B|C) = P(A|C) - P(A|C)P(B|C)$ ■

A and B are not guaranteed to be conditionally independent given C^c. This may be true in some special cases, e.g., if $P(A) = 0$ and $P(B) = 0$. However, it is in general false. Suppose, for example, that events A and B have nonempty intersection inside C, and are conditionally independent, but have empty intersection inside C^c, which would make them dependent (given C^c).

Conditioning may affect Independence

Consider two unfair coins A and B with $P(H|A) = 0.9$, $P(H|B) = 0.1$. Given a particular coin, tosses are independent. Also, coins are equally likely to be chosen. Are overall coin tosses independent?

Consider the particular result H_{11}. Let $H^{10} = H_1 \cap \ldots \cap H_{10}$
Unconditional probability:
$P(H_{11}) = P(A)P(H_{11}|A) + P(B)P(H_{11}|B) = (0.5)(0.9) + (0.5)(0.1) = 0.5$
Conditional probability (given that the first 10 results are H):

$$P(H_{11}|H^{10}) = \frac{P(H_{11} \cap H^{10})}{P(H^{10})}$$
$$P(H_{11}|H^{10}) = \frac{P(A)P(H^{10} \cap H_{11}|A) + P(B)P(H^{10} \cap H_{11}|B)}{P(A)P(H^{10}|A) + P(B)P(H^{10}|B)}$$
$$P(H_{11}|H^{10}) = \frac{(0.9^{11})(0.5) + (0.1^{11})(0.5)}{(0.9^{10})(0.5) + (0.1^{10})(0.5)} \approx 0.9 \Rightarrow \text{Overall tosses aren't independent.}$$

Collective/Mutual Independence

Events $A_1, A_2, \ldots, A_n$ are called independent if:

$$P\left(\bigcap_{i \in S} A_i\right) = \prod_{i \in S} P(A_i), \quad \text{for every subset } S \text{ of } \{1, 2, \ldots, n\}$$

For instance, let $n = 3$, then we'll need the following Pairwise Independence conditions:

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

But we'll also need: $P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3)$, since it doesn't follow from the Pairwise Independence condition.

Furthermore, if the boxed condition holds, then any relation such as:
$P(A_3) = P(A_3|A_1 \cap A_2) = P(A_3|A_1 \cap A_2^c) = P(A_3|A_1^c \cap A_2) \ldots$ also holds.

For a **very detailed and long** sequence of proofs see:
https://stats.libretexts.org/Bookshelves/Probability_Theory/Probability_Mathematical_Statistics_and_Stochastic_Processes_(Siegrist)/02%3A_Probability_Spaces/2.05%3A_Independence
and
https://math.stackexchange.com/questions/1889772/understanding-the-definition-for-collections-of-events-being-independent
Add the proofs to the appendix!

Let A, B, C, D be independent events.
a) Is it guaranteed that $A \cap C$ is independent from $B^c \cap D$?
$A \cap C$ only gives information about A and C, and nothing about D or B (and therefore B^c).

b) Is it guaranteed that $A \cap B^c \cap D$ is independent from $B^c \cap D^c$?
If $A \cap B^c \cap D$ occurs $\Rightarrow$ D occurs $\Rightarrow$ D^c doesn't occur, and so it affects the probability of $B^c \cap D^c$. Also, consider that if $B^c \cap D \neq \emptyset \Rightarrow$ It's a subset of $B^c \cup D^c$ and they have common elements, but we're not considering the intersection with A that could exclude it.

Pairwise Independent, but not Collectively Independent

Consider two independent fair coin tosses $\Rightarrow$ Outcomes in {HH, HT, TH, TT} have the same probability.
Let $\{H_1 = \text{First toss is } H\}$, $\{H_2 = \text{Seconds toss is } H\}$ $\Rightarrow P(H_1) = P(H_2) = 1/2$
Let C : The two tosses have the same result.

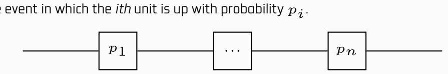
There's pairwise independence between all pairs:
$P(H_1 \cap C) = P(H_1 \cap H_2) = 1/4 \wedge P(H_1)P(C) = (1/2)(1/2) = 1/4$
$P(H_2 \cap C) = P(H_2 \cap H_1) = 1/4 \wedge P(H_2)P(C) = (1/2)(1/2) = 1/4$

But there's no independence for the conjunction of all three events:
$P(H_1 \cap H_2 \cap C) = P(H_1 \cap H_2) = (1/2)^2 = 1/4$
$P(H_1)P(H_2)P(C) = (1/2)^3 = 1/8$

Intuitively, notice that for $P(C|H_1)$, the occurrence of H_1 doesn't give any new information, since the second toss could still be T or H with equal probability, so $P(C|H_1) = (1/2) = P(C)$, but if we are asked for $P(C|H_1 \cap H_2)$ we can see that the answer is 1.

Reliability: Serial Configuration

Let U_i be the event in which the *ith* unit is up with probability p_i.

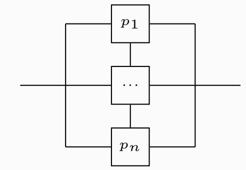


Events U_i are independent, so the probability that the whole serial system is up becomes:

$$\boxed{P\left(\bigcap U_i\right) = \prod p_i}$$

Reliability: Parallel Configuration

Notice that the system is up if any subset (of the set of all components) is up. Let F_i be the event in which the *ith* component fails (these events are also independent), then:



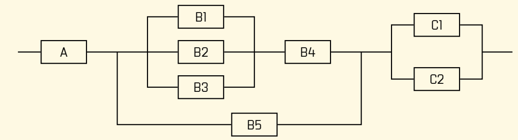
The probability that the whole system is up becomes $P\left(\bigcup U_i\right)$, which can be seen as the complementary probability that all components fail together:

$$1 - P\left(\bigcap F_i\right) = 1 - \prod P(F_i)$$

$$\boxed{1 - \prod (1 - p_i)}$$

Reliability: Complex Configuration

Assume each component has a probability p of *being up*.



The system can be seen as a serial configuration of 3 major groups: A, B, and C. A has probability p of being up, C has probability $1 - (1 - p)^2$ of being up. Now, B is a parallel configuration of 2 subgroups:

- B1, B2, B3 with probability $1 - (1 - p)^3$ in a serial configuration with B4 (probability p), therefore we have: $[1 - (1 - p)^3]p$
- B5 with probability p

And so, for the group B we have: $1 - (1 - [1 - (1 - p)^3 p])(1 - p)$. Finally, this is multiplied by the probabilities of groups A and C.

Counting

Discrete Uniform Probability Law

If Ω consists of n equally likely elements $(P(\omega) = 1/n)$ and A consists of k elements, then

$$\boxed{P(A) = \frac{k}{n}}$$

Basic Counting Principles

If there are r stages and n_i choices at stage i, then the number of choices is:

$$\boxed{n_1 n_2 \ldots n_r = \prod_{i=1}^{r} n_i}$$

Number of subsets of $\{1, \ldots, n\}$

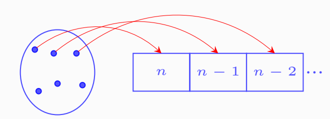
1	2	3	...	n
Y	Y	Y	...	Y
N	N	N	...	N

Consider the set $\{1, \ldots, n\}$, then all subsets (including $\emptyset$) can be formed by making consecutive binary decisions to include (Y) or not include (N) an element.

$\boxed{\text{Number of subsets} = 2^n}$

Permutations

Number of ways of ordering n elements.



$\boxed{n(n-1)(n-2)\ldots(1) = n!}$

You're given the set of letters {A,B,C,D,E}. What is the probability that in a random five-letter string (each letter appears exactly once and all such strings are equally likely) the letters A and B are next to each other?

By fixing the pair "AB" we have 4 elements in the set $\Rightarrow$ Probability = $\frac{4! \times 2}{5!}$

Combinations

Let $\binom{n}{k}$ be the total number of **unique combinations** of k elements from a set of n elements. To obtain this number we can proceed with:

$$\underbrace{n(n-1)(n-2)\ldots(n-k+1)}_{k \text{ elements}} = \frac{n!}{(n-k)!}$$

But by doing so, we are counting repeated subsets such as {A,B,C,D,E} and {B,A,C,D,E}. So, in order to avoid permutations, we divide by $k!$, therefore:

$$\boxed{\binom{n}{k} = \frac{n!}{k!(n-k)!}}$$

Binomial Theorem

Notice that in the expansion of:

$$(x+y)^n = \underbrace{(x+y)\ldots(x+y)}_{k \text{ factors}}\underbrace{(x+y)\ldots(x+y)}_{n-k \text{ factors}}$$

there are $\binom{n}{k}$ elements of the form $x^k y^{n-k}$. So, in order to account for the grouped sum of all terms:

$$\boxed{(x+y)^n = \sum_{k=0}^{n}\binom{n}{k}x^k y^{n-k} \;\blacksquare}$$

Another approach for the number of subsets

Notice that: $\sum_{k=0}^{n}\binom{n}{k} = \binom{n}{0}+\binom{n}{1}+\ldots+\binom{n}{n}$ = Number of subsets, therefore:

$$\boxed{\sum_{k=0}^{n}\binom{n}{k} = (1+1)^n = 2^n}$$

Partitions

We have n elements and we need to form r groups of n_i elements each such that $\sum n_i = n$.
Let the following expression represent that idea:

$$M = \binom{n}{n_1, n_2, \ldots, n_r}$$

Notice that if we first take n_1 elements, we'd have $(n-n_1)$ from which we can take n_2, and so on...

$$M = \binom{n}{n_1}\binom{n-n_1}{n_2}\binom{n-n_1-n_2}{n_3}\ldots\binom{n-n_1-\ldots-n_{r-1}}{n_r}$$

$$M = \frac{n!}{n_1!(n-n_1)!}\cdot\frac{(n-n_1)!}{n_2!(n-n_1-n_2)!}\cdot\ldots\cdot\frac{(n-n_1-n_2-\ldots-n_{r-1})!}{n_r!(0!)}$$

Then

$$\boxed{\binom{n}{n_1, n_2, \ldots, n_r} = \frac{n!}{n_1!n_2!\ldots n_r!} \;\blacksquare}$$

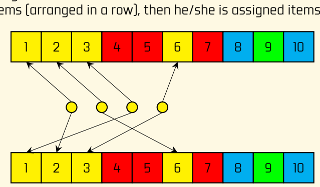
Chess Tournament & Item Distribution

Consider the following two problems:

- **Chess tournament:** There are 10 competitors: 4 Russian, 3 from US, 2 from UK, 1 from Brazil. If the tournament result lists just the nationalities of the players, how many outcomes are possible?
- **Distribution of items:** 10 items are to be distributed to 4 people such that they receive 4, 3, 2, 1 items, respectively, how many outcomes are possible?

The answer for both problems is: $\frac{10!}{4!3!2!1!}$. What is the underlyig logic the relates them?

Imagine that the item distribution is executed in the following fashion: Person 1 is placed in front of the items (arranged in a row), then he/she is assigned items 1,2,3,4 or 1,2,4,3 or ... (i.e. the outcome is the same).

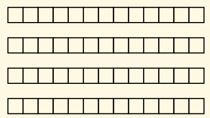


Now imagine that the yellow balls represent the four russian players and also the four item assignments to person 1, for the chess and item distribution problems respectively.

It doesn't matter which yellow ball is assigned first, different permutations for this particular arrangement/outcome are counted as one.

Ace distribution problem

There's a 52-card deck, dealt (fairly) to four players. Find the probability that each player gets an ace.



Combinatorial Approach: There are $4!$ ways of distributing the aces. For the remaining cards, there are $\frac{48!}{(12!)^4}$ possibilities.

$|\Omega| = \frac{52!}{(13!)^4} \Rightarrow P(\text{Event}) = 4!\,\frac{48!}{(12!)^4} / \frac{52!}{(13!)^4}$

Conditional Probability Approach: After we randomly deliver the first ace to one of the players, there are 39 slots out of the remaining 51 which do not correspond to the player that got an ace.
Following this logic $\Rightarrow P(\text{Event}) = (\frac{52}{52})(\frac{39}{51})(\frac{26}{50})(\frac{13}{49})$

Multinomial Theorem

The expansion of $(x_1+x_2+\ldots+x_r)^n$ will produce r^n elements of the form $x_1^{n_1}x_2^{n_2}\ldots x_r^{n_r}$ such that $\sum n_i = n$.

Some of these elements are identical and can be grouped exactly as those from the expansion in the Binomial Theorem.

This is equivalent to listing all the possible divisions of n distinct elements into r distinct groups of sizes $n_1, n_2, \ldots, n_r$, but there are more than one such combinations that add up to n, so:

$$\boxed{(x_1+x_2+\ldots+x_r)^n = \sum_{\substack{(n_1,\ldots,n_r):\\ n_1+\ldots+n_r=n}}\binom{n}{n_1,n_2,\ldots,n_r}x_1^{n_1}x_2^{n_2}\ldots x_r^{n_r}}$$

Counting committees

We start with a pool n of people. A chaired committee consists of $k \geq 1$ members, out of whom one member is designated as the chairperson. Then, the total number of possible chaired committees of any size is: $\sum_{k=1}^{n} k\binom{n}{k}$. On the other hand, we can also get to the result by first selecting a chairperson and then the committee: n possibilities for the chairperson, 2^{n-1} possible subsets of $n-1$ (the empty set + the chairperson would make a 1-man committee), so $\sum_{k=1}^{n} k\binom{n}{k} = n2^{n-1}$.

Binomial Probabilities

Consider an experiment with a binary outcome (e.g. success/failure, yes/no, heads/tails, 0/1).
Let p be the probability of success $\Rightarrow$ The probability of k successes in n **independent** trials is:

$$\boxed{P(k \text{ successes}) = \binom{n}{k}p^k(1-p)^{n-k}}$$

Also, notice that $\sum_{k=0}^{n}\binom{n}{k}p^k(1-p)^{n-k} = 1$

Conditional coin tossing

Find the probability that the 6th toss out of a total of 10 tosses is Heads, given that there are exactly 2 Heads out of the 10 tosses.

Binomial Probabilistic Approach:

$P(B) = \binom{10}{2}p^2(1-p)^8,\;\; P(A\cap B) = P(H_6 \cap \text{One more } H) = p\binom{9}{1}p(1-p)^8$

$\Rightarrow P(A|B) = \binom{9}{1}/\binom{10}{2}$

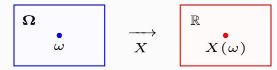
Conditional Uniformity Approach:

- Consider Ω as the set of all possible sequences of 10-tosses.
- Let B be the subset that includes exactly 2 heads. There are $\binom{10}{2}$ such elements, all with probability $p^2(1-p)^8$.
- Let A be the subset that includes H_6. Since H_6 is fixed, there are $\binom{9}{1}$ ways for $A\cap B$ to occur. And because these elements are included in B, they have probability $p^2(1-p)^8$.
- $\Rightarrow P(A|B) = \binom{9}{1}/\binom{10}{2}$

Discrete Random Variables

Random Variables

Random Variable



Simple Definition:
A random variable $X(\omega)$ or simply X is a function defined on the sample space Ω that associates its outcomes to $\mathbb{R}$.

$$X : \Omega \to \mathbb{R}$$

- The term "random variable" can be misleading since it's a function rather than a variable.
- More formally, it is a measurable function $X : \Omega \to R$ from Ω into a measurable space R.
- The outcome itself can be thought of as a random variable
  $\Rightarrow \Omega = R,\, X(\omega) = \omega \to$ Identity function.

Notation

- $X(\omega)$ or simply X is a random variable, a function defined on the domain Ω with range in $\mathbb{R}$.
- x is an unspecified value of $X \Rightarrow x$ is a real variable such that $x \in \mathbb{R}$.
- $\{X = a\} \Longleftrightarrow \{\omega : X(\omega) = a\}$, both imply an **event** in which X takes the particular value a.
- $\{X = a\} = \{\omega : X(\omega) = a\}$ only if the outcome itself is the random variable.
- $\{X = x\}$ is an unspecified (variable) event in which X takes the unspecified value x.
- $P(\{X = x\})$ is the probability of that unspecified event $\to$ We will write $P(X = x)$ for short.
  $\Rightarrow P(\{\omega : X(\omega) = 3\}) \Longleftrightarrow P(X = 3)$

Function of Random Variables

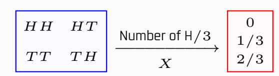
> Let X be a random variable with values in $S \Rightarrow X : \Omega \to S \subseteq \mathbb{R}$.
> Let g be a function defined on S that maps into $T \Rightarrow g : S \to T$.
> $\Rightarrow g(X)$ depends on $X \Rightarrow g(X)$ depends on the values in Ω.
> $\Rightarrow Y = g(X)$ is a random variable with values in T. $\blacksquare$

> Let X and Y be random variables.
> Let $g(X,Y)$ be a function of the random variables X and Y.
> $\forall x, y : (X = x \wedge Y = y) \Rightarrow g(X,Y) = g(x,y)$.
> However, the converse is not necessarily true.

Let $X \in \{1,2\} \wedge Y \in \{3,4\}$
So $X = 2 \wedge Y = 3 \Rightarrow X + Y = 5$
but $X + Y = 5 \Rightarrow (X = 2 \wedge Y = 3) \vee (X = 1 \wedge Y = 4)$ $\blacksquare$

Discrete Random Variable

> A random variable is called discrete if its **range** is either finite or countably infinite.



Experiment: 2 tosses of a coin.
Let X be the number of heads divided by 3.
The range is finite: consists of 3 elements.

$$\text{sgn}(\omega) = \begin{cases} -1, & \omega < 0 \\ 0, & \omega = 0 \\ 1, & \omega > 0 \end{cases}$$

Experiment: Select $\omega : \omega \in [-1, 1]$
Let $X = \text{sgn}(\omega)$
The domain is infinite, but the range is finite.

Probability Mass Functions & Expectation

Probability Mass Function

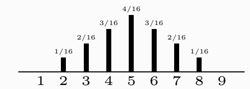
> The function
> $$p_X(x) = P(X = x) = P(\omega : X(\omega) = x)$$
> is called the Probability Mass Function (**PMF**) and assigns a probability to each numerical value of a Discrete Random Variable.

For simplicity, let $S = \Omega$ be the sample space. The PMF has the following properties:

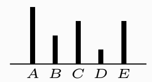
- $p_X(x) \geq 0,\, x \in S$
- $\sum_{x\in S} p_X(x) = 1$
- $\sum_{x\in A} p_X(x) = P(A),\;\; A \subseteq S$

Graphs of PMF

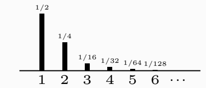
Sum of 2 rolls of a tetrahedral die.



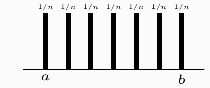
PMF of a categorical random variable.



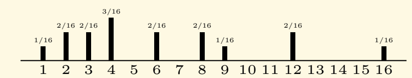
$P(X = x) = \frac{1}{2^x}$, for $x \in \mathbb{N}$.



Discrete Uniform Random Variable.



Consider the experiment of tossing twice a fair tetrahedral die. Let X be the product of the rolls, then:



×	1	2	3	4
1	1	2	3	4
2	2	4	6	8
3	3	6	9	12
4	4	8	12	16

$p_X(4) = P(\{1,4\} \cup \{2,2\} \cup \{4,1\}) = 3/16$
$p_X(5) = 0$

EOQ formula derivation

Since demand is deterministic, we can get rid of the Stockout Cost concept for now. So,

$$TRC(Q) = c_t \frac{D}{Q} + c_e \frac{Q}{2}$$

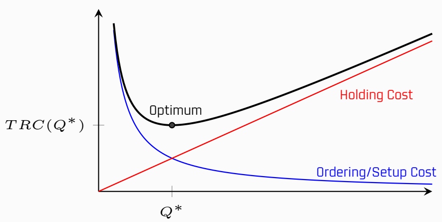
From the first-order optimal condition (first derivative equals zero), we have

$$0 = \frac{\mathrm{d}}{\mathrm{d}Q}\left(\frac{c_t D}{Q}\right) + \frac{\mathrm{d}}{\mathrm{d}Q}\left(\frac{c_e Q}{2}\right)$$

$$0 = -\frac{c_t D}{Q^2} + \frac{c_e}{2}$$

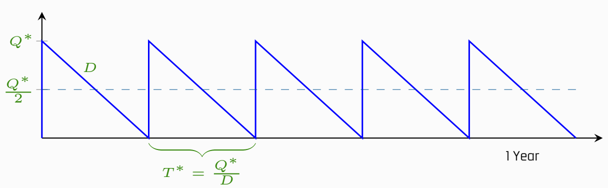
$$\boxed{Q^* = \sqrt{\frac{2c_t D}{c_e}}}$$

The EOQ or Q^* gives the minimum TRC under deterministic conditions:



EOQ sawtooth plot

The optimal policy becomes ordering Q^* units of inventory every T^* units of time.



Notice that the total consumption of the last order may take place after the 1 year (unit time) period.

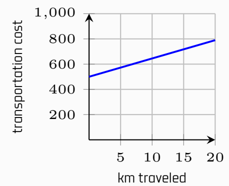
Mathematical Functions

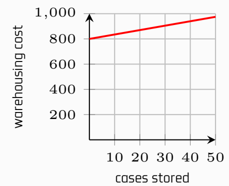
Linear Functions

$$f(x) = mx + b$$

Cost functions:

f(Level of Activity) $=$ Fixed Cost $+$ Variable Cost(Level of Activity)





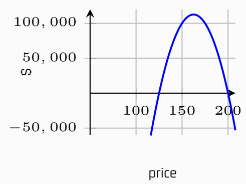
Linear Regressions

fig

Quadratic Functions

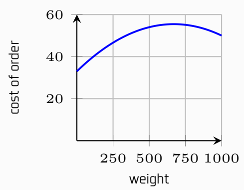
$$f(x) = ax^2 + bx + c$$

Profit:



$V(p) = 20,000 - 80p$
$R(p) = (20,000 - 80p)p$
$C(p) = 500,000 + 75(20,000 - 80p)$
$P(p) = R(p) - C(p)$

Parcel trucking



$f(w) = 33 + 0.067w - 0.00005w^2$

Proofs:

Inclusion-Exclusion Principle

Consider the cases for $n=3$ and $n=4$:

$$
\begin{aligned}
P(A_1 \cup A_2 \cup A_3) = \;& P(A_1) + P(A_2) + P(A_3) \\
& - \underbrace{P(A_1 \cap A_2)}_{1<2} - \underbrace{P(A_1 \cap A_3)}_{1<3} - \underbrace{P(A_2 \cap A_3)}_{2<3} \\
& + \underbrace{P(A_1 \cap A_2 \cap A_3)}_{1<2<3}
\end{aligned}
$$

$$
\begin{aligned}
P(A_1 \cup A_2 \cup A_3 \cup A_4) = \;& P(A_1) + P(A_2) + P(A_3) + P(A_4) \\
& - \underbrace{P(A_1 \cap A_2)}_{1<2} - \underbrace{P(A_1 \cap A_3)}_{1<3} - \underbrace{P(A_1 \cap A_4)}_{1<4} - \underbrace{P(A_2 \cap A_3)}_{2<3} - \underbrace{P(A_2 \cap A_4)}_{2<4} - \underbrace{P(A_3 \cap A_4)}_{3<4} \\
& + \underbrace{P(A_1 \cap A_2 \cap A_3)}_{1<2<3} + \underbrace{P(A_1 \cap A_2 \cap A_4)}_{1<2<4} + \underbrace{P(A_1 \cap A_3 \cap A_4)}_{1<3<4} + \underbrace{P(A_2 \cap A_3 \cap A_4)}_{2<3<4} \\
& - \underbrace{P(A_1 \cap A_2 \cap A_3 \cap A_4)}_{1<2<3<4}.
\end{aligned}
$$

We argue that we have a general pattern:

$$
\begin{aligned}
P\left(\bigcup_{i=1}^{n} A_i\right) =& -(-1)^1 \sum_{1 \le i \le n} P(A_i) \\
& -(-1)^2 \sum_{1 \le i_1 < i_2 \le n} P(A_{i_1} \cap A_{i_2}) \\
& -(-1)^3 \sum_{1 \le i_1 < i_2 < i_3 \le n} P(A_{i_1} \cap A_{i_2} \cap A_{i_3}) \\
& -(-1)^4 \sum_{1 \le i_1 < i_2 < i_3 < i_4 \le n} P(A_{i_1} \cap A_{i_2} \cap A_{i_3} \cap A_{i_4}) \\
& \vdots \\
& -(-1)^n P(A_1 \cap A_2 \cap A_3 \cap A_4 \cap \ldots \cap A_n)
\end{aligned}
$$

$$
\boxed{P\left(\bigcup_{i=1}^{n} A_i\right) = -\sum_{k=1}^{n} (-1)^k \sum_{1 \le i_1 < \ldots < i_k \le n} P\left(\bigcap_{j=1}^{k} A_{i_j}\right)}
$$

Proof by Induction:
Suppose the pattern is true for n, we need to show it works for $n+1$. First, consider $n=2$ and apply distributivity:

$$
\begin{aligned}
P(A_1 \cup A_2 \cup \ldots \cup A_n \cup A_{n+1}) &= P\big((A_1 \cup A_2 \cup \ldots \cup A_n) \cup A_{n+1}\big) \\
&= P(A_1 \cup A_2 \cup \ldots \cup A_n) + P(A_{n+1}) - P\big((A_1 \cup A_2 \cup \ldots \cup A_n) \cap A_{n+1}\big) \\
&= \underbrace{P(A_1 \cup A_2 \cup \ldots \cup A_n)}_{n \text{ unions}} + P(A_{n+1}) - \underbrace{P\big((A_1 \cap A_{n+1}) \cup (A_2 \cap A_{n+1}) \cup \ldots \cup (A_n \cap A_{n+1})\big)}_{n \text{ unions}}
\end{aligned}
$$

The first and the last terms are n-unions, for which we assumed the formula to hold. Therefore:

$$P(A_1 \cup A_2 \cup \ldots \cup A_n \cup A_{n+1}) = -(-1)^1 \sum_{1 \le i \le n} P(A_i) \tag{1}$$

$$-(-1)^2 \sum_{1 \le i_1 < i_2 \le n} P(A_{i_1} \cap A_{i_2}) \tag{2}$$

$$-(-1)^3 \sum_{1 \le i_1 < i_2 < i_3 \le n} P(A_{i_1} \cap A_{i_2} \cap A_{i_3}) \tag{3}$$

$$- \ldots -(-1)^n P(A_1 \cap A_2 \cap A_3 \cap A_4 \cap \ldots \cap A_n) \tag{4}$$

$$+ P(A_{n+1}) \tag{5}$$

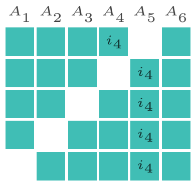
$$+ (-1)^1 \sum_{1 \le i \le n} P(A_i \cap A_{n+1}) \tag{6}$$

$$+ (-1)^2 \sum_{1 \le i_1 < i_2 \le n} P(A_{i_1} \cap A_{i_2} \cap A_{n+1}) \tag{7}$$

$$+ \ldots + (-1)^{n-1} \sum_{1 \le i_1 < i_2 < \ldots < i_{n-1} \le n} P(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_{n-1}} \cap A_{n+1}) \tag{8}$$

$$+ (-1)^n P(A_1 \cap A_2 \cap A_3 \cap A_4 \cap \ldots \cap A_n \cap A_{n+1}) \tag{9}$$

Here (1) and (5) account for all the probabilities of single events from 1 to $n+1$. (2) includes all the two- intersection probabilities from 1 to n, and (6) all the two-intersection probabilities where the higher index equals $n+1$. These two sums thus account for all possible two-intersection probabilities from 1 to $n+1$. Similarly, (3) includes all three-intersection probabilities from 1 to n, and (7) those with highest index equal to $n+1$. Together they include all three-intersection probabilities from 1 to $n+1$.



This continues until (4) and (8), which together give all n-intersection probabilities from 1 to $n+1$. To see why this is true, let's consider the case for $n=5$ (i.e. we would prove that the formula applies for $n=6$). It could be the case that $A_{i_{n-1}} = A_{i_4} = A_5$ (see the figure), so equation (8) would give all the combinations on the figure (emerald squares), and equation number (4) would give the missing intersection: $A_1 \cap A_2 \cap A_3 \cap A_4 \cap A_5$.

Finally, we write the last term (9) and, therefore, we observe that:

$$
\begin{aligned}
P\left(\bigcup_{i=1}^{n+1} A_i\right) =& -(-1)^1 \sum_{1 \le i \le n+1} P(A_i) \\
& -(-1)^2 \sum_{1 \le i_1 < i_2 \le n+1} P(A_{i_1} \cap A_{i_2}) \\
& -(-1)^3 \sum_{1 \le i_1 < i_2 < i_3 \le n+1} P(A_{i_1} \cap A_{i_2} \cap A_{i_3}) \\
& - \ldots -(-1)^n \sum_{1 \le i_1 < i_2 < \ldots < i_n \le n+1} P(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_n}) \\
& -(-1)^{n+1} P(A_1 \cap A_2 \cap A_3 \cap A_4 \cap \ldots \cap A_{n+1})
\end{aligned}
$$

We have proven that the expression works for $n+1$ ■

References:

Books

2008 - Introduction to Probability (2nd ed.) - Dimitri P. Bertsekas & John N. Tsitsiklis
2021 - Probability, Mathematical Statistics, and Stochastic Processes - Kyle Siegrist

MIT OpenCourseWare

https://www.youtube.com/playlist?list=PLUl4u3cNGP60A3XMwZ5sep719_nh95qOe

Links

Inclusion-Exclusion Principle
https://math.stackexchange.com/questions/2587979/generalized-formula-for-the-probability-of-the-union-of-n-events-occurring
https://people.maths.bris.ac.uk/~mb13434/incl_excl_n.pdf
Event Independence
https://math.stackexchange.com/questions/1832686/probability-are-disjoint-events-independent
Random Variables
https://en.wikipedia.org/wiki/Random_variable
Siegrist: Random Variables