ITC8190
Mathematics for Computer Science
Elementary Probability Theory

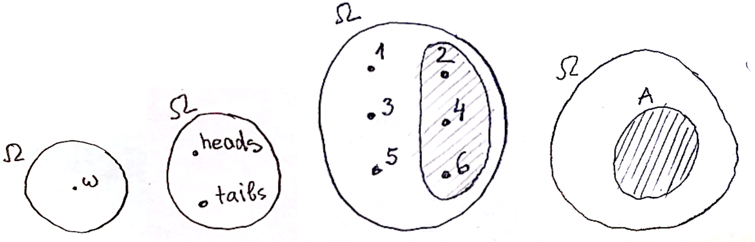
Aleksandr Lenin

November 27th, 2018

The following slides were borrowed from a lecture material in cryptography course by prof. Ahto Buldas with his permission.

`https://courses.cs.ttu.ee/w/images/c/c7/ITC8240-Unbreakable-ciphers.pdf`

Ω-sample space, that contains all possible outcomes $\omega \in \Omega$.

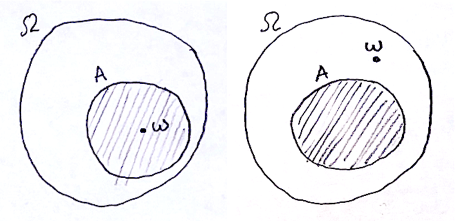


For example, $\Omega = \{\text{heads}, \text{tails}\}$ for a coin, and $\Omega = \{1, \ldots, 6\}$ for a die.

Events are subsets $A \subseteq \Omega$.

For a die, the event $\{2, 4, 6\}$ means that the outcome is even.

An event A *happens* if $\omega \in A$ for the actual outcome ω.

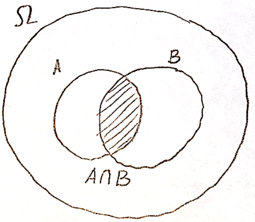


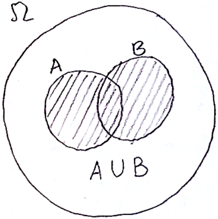
Empty event $\emptyset$ is called the *impossible event* (it *never* happens)

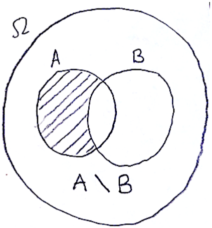
Ω is called the *universal event* (it *always* happens)

For every two events A and B we can compute:

Intersection	*A and B*	$A \cap B = \{\omega \in \Omega \colon \omega \in A \text{ and } \omega \in B\}$
Union	*A or B*	$A \cup B = \{\omega \in \Omega \colon \omega \in A \text{ or } \omega \in B\}$
Difference	*A but not B*	$A \backslash B = \{\omega \in \Omega \colon \omega \in A \text{ and } \omega \notin B\}$

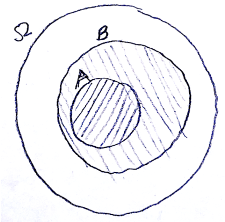


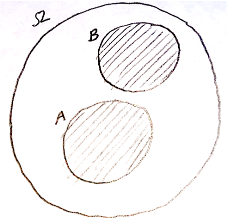




Inclusion: Event A *implies* event A, if $A \subseteq B$, i.e. if $\omega \in A$ always implies $\omega \in B$. If A happens then B happens.

Exclusion: Events A and B are *mutually exclusive* if $A \cap B = \emptyset$, i.e. A and B cannot simultaneously happen.





Theorem (1)

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

Proof.

We prove (a) $A \subseteq (A \backslash B) \cup (A \cap B)$ and (b) $(A \backslash B) \cup (A \cap B) \subseteq A$

(a) If $\omega \in A$ then either:

- $\omega \in B$, which implies $\omega \in A \cap B$, or
- $\omega \notin B$, which implies $\omega \in A \backslash B$

(b) If $\omega \in (A \backslash B) \cup (A \cap B)$, then either:

- $\omega \in A \backslash B$, which implies $\omega \in A$, or
- $\omega \in A \cap B$, which also implies $\omega \in A$ □

Theorem (2)

$A \cup B = (A \backslash B) \cup B$

Proof.

We prove (a) $A \cup B \subseteq (A \backslash B) \cup B$ and (b) $(A \backslash B) \cup B \subseteq A \cup B$

(a) If $\omega \in A \cup B$, then either:

- $\omega \in B$ or
- $\omega \notin B$ and $\omega \in A$, which implies $\omega \in A \backslash B$.

(b) If $\omega \in (A \backslash B) \cup B$ then either:

- $\omega \in B$ or
- $\omega \in A \backslash B$ that implies $\omega \in A$. □

The set $\mathcal{F}$ of all events we consider must be a *sigma-algebra*:

- $\Omega \in \mathcal{F}$
- If $A \in \mathcal{F}$, then $\Omega \backslash A \in F$
- If $A_1, A_2, A_3, \ldots \in \mathcal{F}$, then $A_1 \cup A_2 \cup A_3 \cup \ldots \in \mathcal{F}$

If $A \in \mathcal{F}$, then A is said to be a *measurable* subset.

Example: The *set $P(\Omega)$ of all subsets* of Ω is a sigma-algebra.

In this class, we mostly assume that $\mathcal{F} = P(\Omega)$.

Probability (measure) is a function $\mathsf{P}\colon \mathcal{F} \to \mathbb{R}$ such that:

- *PM1:* $0 \leq \mathsf{P}[A] \leq 1$ for any event $A \in \mathcal{F}$.
- *PM2:* $\mathsf{P}[\Omega] = 1$
- *PM3:* If $A_1, A_2, \ldots \in \mathcal{F}$ are mutually exclusive, then

$$\mathsf{P}[A_1 \cup A_2 \cup \ldots] = \mathsf{P}[A_1] + \mathsf{P}[A_2] + \ldots$$

The triple $(\Omega, \mathcal{F}, \mathsf{P})$ is called a *probability space*.

If $\mathcal{F}$ is the set of all subsets of Ω, we omit $\mathcal{F}$ and say that a probability space is a pair (Ω, P).

Theorem

$\mathsf{P}[\Omega \backslash A] = 1 - \mathsf{P}[A]$

Proof.

By *PM2*, we have $\mathsf{P}[\Omega] = 1$. As A and $\Omega \backslash A$ are mutually exclusive, and $(\Omega \backslash A) \cup A = \Omega$, by *PM3*, we have $\mathsf{P}[\Omega \backslash A] + \mathsf{P}[A] = \mathsf{P}[\Omega] = 1$ and hence

$$\mathsf{P}[\Omega \backslash A] = \underbrace{\mathsf{P}[\Omega \backslash A] + \mathsf{P}[A]}_{1} - \mathsf{P}[A] = 1 - \mathsf{P}[A] \ .$$

□

Theorem

$\mathsf{P}[A] + \mathsf{P}[B] = \mathsf{P}[A \cap B] + \mathsf{P}[A \cup B]$

Proof.

By Thm. 1: $A = (A \backslash B) \cup (A \cap B)$. As $A \backslash B$ and $A \cap B$ are mutually exclusive, by *PM3*: $\mathsf{P}[A] = \mathsf{P}[A \backslash B] + \mathsf{P}[A \cap B]$. Hence,

$$\mathsf{P}[A] + \mathsf{P}[B] = \mathsf{P}[A \backslash B] + \mathsf{P}[B] + \mathsf{P}[A \cap B]$$

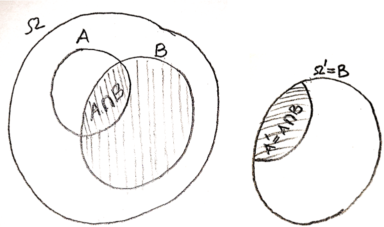
By Thm. 2: $A \cup B = (A \backslash B) \cup B$. As $A \backslash B$ and B are mutually exclusive, by *PM3*: $\mathsf{P}[A \cup B] = \mathsf{P}[A \backslash B] + \mathsf{P}[B]$. Hence,

$$\mathsf{P}[A] + \mathsf{P}[B] = \underbrace{\mathsf{P}[A \backslash B] + \mathsf{P}[B]}_{\mathsf{P}[A \cup B]} + \mathsf{P}[A \cap B] = \mathsf{P}[A \cup B] + \mathsf{P}[A \cap B] \ .$$

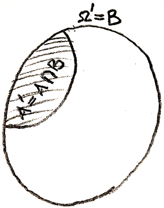
□

Somehow we learn that an event B (with $\mathsf{P}[B] \neq 0$) happens, i.e. $\omega \in B$.

Probability space (Ω, P) collapses to a new space (Ω', P'), where $\Omega' = B$.



Magnify by β



We want that there is β, so that $\mathsf{P}'[A] = \beta \cdot \mathsf{P}[A \cap B]$ for any event A.

As in the new space, $\mathsf{P}'[B] = \mathsf{P}'[\Omega'] = 1$, we have $\beta = \frac{1}{\mathsf{P}[B \cap B]} = \frac{1}{\mathsf{P}[B]}$, i.e.

$$\mathsf{P}'[A] = \frac{\mathsf{P}[A \cap B]}{\mathsf{P}[B]} \quad .$$

The probability

$$P'[A] = \frac{P[A \cap B]}{P[B]}$$

is denoted by $P[A \mid B]$ and is called the *conditional probability* of A assuming that B happens, i.e.

$$P[A \mid B] = \frac{P[A \cap B]}{P[B]}$$

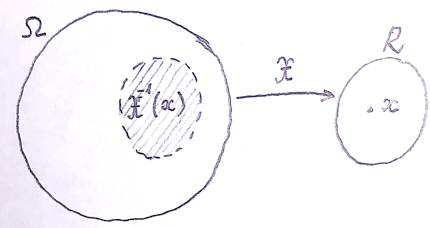
Corollary (Chain Rule):

$$P[A \cap B] = P[B] \cdot P[A|B] = P[A] \cdot P[B|A]$$

Random variable X is any function $X\colon \Omega \to R$, where R is called the *range* of X. We write R_X to denote the range of X

For any $x \in R$, we define $X^{-1}(x)$ as the event $\{\omega\colon X(\omega) = x\}$ and use the notation:

$$\underset{X}{\mathsf{P}}[x] = \mathsf{P}[X = x] = \mathsf{P}[X^{-1}(x)] \ .$$



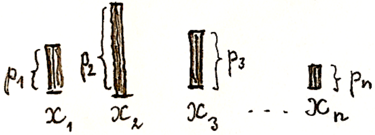
In cryptography, we mostly assume that the range R is *finite*.

Note that if $x \neq x'$, then the events $X^{-1}(x)$ and $X^{-1}(x')$ are mutually exclusive and as $\cup_{x \in R} X^{-1}(x) = \Omega$, we have:

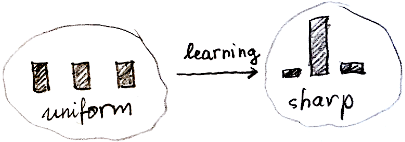
$$\sum_{x} \underset{X}{\mathsf{P}}[x] = \mathsf{P}[\cup_{x \in R} X^{-1}(x)] = \mathsf{P}[\Omega] = 1 \ .$$

Assume R is finite and $R = \{x_1, x_2, \ldots, x_n\}$.

The sequence of values $(p_1, p_2, \ldots, p_n)$, where $p_i = \underset{X}{\mathsf{P}}[x_i]$, is called the *probability distribution* of X.



Histograms are graphical representations of probability distributions.



Events A and B are said to be *independent* if $\mathsf{P}[A \cap B] = \mathsf{P}[A] \cdot \mathsf{P}[B]$

If $\mathsf{P}[A] \neq 0 \neq \mathsf{P}[B]$, then independence is equivalent to:

$$\mathsf{P}[A \mid B] = \mathsf{P}[A] \qquad \text{and} \qquad \mathsf{P}[B \mid A] = \mathsf{P}[B] \ ,$$

i.e. the probability of A does not change, if we learn that B happened.

We say that X and Y are *independent random variables* if for every $x \in R_X$ and $y \in R_Y$:

$$\begin{aligned} \mathsf{P}[X = x, Y = y] &= \mathsf{P}[X^{-1}(x) \cap Y^{-1}(y)] = \mathsf{P}[X^{-1}(x)] \cdot \mathsf{P}[Y^{-1}(y)] \\ &= \mathsf{P}[X = x] \cdot \mathsf{P}[Y = y] \ . \end{aligned}$$

This means that the probability distribution of X does not change, if we learn the value of Y, and vice versa

By the *direct product* XY (or (X, Y)) of random variables X and Y on a probability space (Ω, P) is a random variable defined by

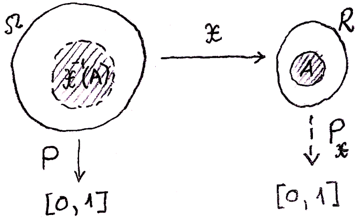
$$(XY)(\omega) = (X(\omega), Y(\omega)) \ .$$

Let X be a random variable (with range R) on a probability space (Ω, P).

If we take $\Omega' = R$ and define a probability function $\underset{X}{\mathsf{P}}$ on R as follows:

$$\underset{X}{\mathsf{P}}[A] = \mathsf{P}[X^{-1}(A)]$$

where $X^{-1}(A) = \{\omega \in \Omega \colon X(\omega) \in A\}$, we get a probability space $(R, \underset{X}{\mathsf{P}})$ that we call a *factor space*.



To sum up, the chain rule is

$$\Pr[A \cap B] = \Pr[A|B] \cdot \Pr[B] = \Pr[B|A] \cdot \Pr[A] \ .$$

If events A and B are **independent**, then $\Pr[A|B] = \Pr[A]$ and $\Pr[B|A] = \Pr[B]$, and the chain rule takes the form of

$$\Pr[A \cap B] = \Pr[A] \cdot \Pr[B] \ .$$

The probability of the union

$$\Pr[A \cup B] = \Pr[A] + \Pr[B] - \Pr[A \cap B] \ .$$

If events A and B are **mutually exclusive**, then $\Pr[A \cap B] = 0$ and hence

$$\Pr[A \cup B] = \Pr[A] + \Pr[B] \ .$$

The chain rule

$$\Pr[A \cap B] = \Pr[A|B] \cdot \Pr[B] = \Pr[B|A] \cdot \Pr[A] \ .$$

also provides us with the relationship between conditional probabilities $\Pr[A|B]$ and $\Pr[B|A]$, namely

$$\Pr[A|B] = \frac{\Pr[B|A] \cdot \Pr[A]}{\Pr[B]} \ ,$$

where:

$\Pr[A]$ is the prior belief
$\Pr[B|A]$ is called the likelihood
$\Pr[B]$ is called evidence
$\Pr[A|B]$ is called the posterior

This is known as the **Bayes' theorem**. It allows to make informed guesses about observations based on prior knowledge or beliefs.

THANK YOU
FOR
YOUR
ATTENTION
ANY QUESTIONS?