Suppose we perform an experiment. For example

~ Roll a dice 100 times

data = list of 100 numbers in $\{1,2,3,4,5,6\}$.

What do we expect the data set to look like?

What is the chance (probability, liklihood) that the dataset has all odd numbers $\{1,3,5\}$?

A probability is a measure which assigns a value in $[0,1]$ to an event A.

eg A = event where 100 rolls of dice always gives an odd number.

$\mathbb{P}(A) = 1/8$ gives a weight or way of measuring the liklihood of an event.

$\boxed{\text{Definition}}$ A probability space is a triple $(\Omega, \mathcal{A}, \mathbb{P})$

- Ω is a nonempty set (sample space)
- $\mathcal{A} \subseteq 2^{\Omega} = \left\{ \begin{matrix}\text{set of all} \\ \text{subsets of } \Omega\end{matrix} \right\}$ is a σ-algebra*
- $\mathbb{P} : \mathcal{A} \longrightarrow [0,1]$ is a probability measure.

i.e. $\mathbb{P}(\Omega) = 1$, $\mathbb{P}\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n\in\mathbb{N}} \mathbb{P}(A_n)$ when $A_i \cap A_j = \emptyset \ \forall i \neq j$.

* Definition of the σ-algebra

$\mathcal{A}$ must satisfy the following

- $\Omega \in \mathcal{A}$, $\emptyset \in \mathcal{A}$
- Closed under complements $[B \in \mathcal{A} \Rightarrow \Omega \setminus B \in \mathcal{A}]$
- Closed under countable unions $\left[\begin{array}{l} B_i \in \mathcal{A} \ \forall i \in \mathbb{N} \\ \Rightarrow \bigcup_{i \in \mathbb{N}} B_i \in \mathcal{A} \end{array}\right]$

Every set $A \in \mathcal{A}$ is called an <u>event</u>

Example 1 First two σ-algebras

(a) $\{\emptyset, \Omega\}$ is a σ-algebra

(b) 2^{Ω} = set of all subsets of Ω is a σ-algebra.

Example 2 Rolling dice

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

$\mathcal{A} = 2^{\Omega} = \{\emptyset, \Omega, \{1\}, \{1, 2, 3, 4, 5\}, \dots\}$

Size of $\mathcal{A} = 2^6 = 64$, so there are 64 possible events.

$\boxed{\text{Example 3}}$ Flip a coin twice

$$\Omega = \{ \{TT\}, \{HH\}, \{TH\}, \{HT\} \}$$

$\mathcal{A} = 2^{\Omega}$ size of $\mathcal{A} = 2^{|\Omega|} = 2^4 = 16.$

$\boxed{\text{Remark}}$ When Ω is discrete or finite, we can always take $\mathcal{A} = 2^{\Omega}$ as the σ-algebra.

$\boxed{\text{Example 4}}$ $\Omega = \mathbb{R}$

When Ω is not discrete, we cannot take 2^{Ω}.

(by measure theory, but basically because 2^{Ω} is too large)

Instead we set

$\mathcal{A}$ = the Borel σ-algebra

what we need : $\mathcal{A}$ contains all intervals $(a, b]$, $a<b$.

An element $A \in \mathcal{A}$ is called a Borel set.

Bayes Theorem We fix a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Let $A, B \in \mathcal{A}$. Sometimes we want to describe the probability an event A will occur given the fact that another event B has happened. This is denoted $\mathbb{P}(A|B)$ and read as "the probability of A given B".

Def Assume $\mathbb{P}(B) > 0$. The <u>conditional probability</u> of A given B is

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

ex $\Omega = \{1, 2, 3, 4, 5, 6\}$. $A = \{2\}$, $B = \{2, 4, 6\}$

$\mathbb{P}(\{k\}) = \frac{1}{6}$ for $k = 1, \ldots, 6$. $\mathbb{P}(A) = 1/6$.

However if we know only even #s will be drawn,

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}.$$

$\boxed{\text{Bayes' Theorem}}$ Let $A, B \in \mathcal{A}$; $\mathbb{P}(A), \mathbb{P}(B) > 0$.

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

$\boxed{\text{Pf}}$

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

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

$\square$

Random Variables

Recall when not working with discrete variables, we have to work with the Borel σ-algebra instead of 2^{Ω}. We want to define functions which behave well with the σ-algebra

$\boxed{\text{Def}}$ Let $(X, \mathcal{A}), (Y, \mathcal{B})$ be probability spaces.

A function $f: X \to Y$ is called <u>measurable</u>

if $\forall A \in \mathcal{B}$

$$f^{-1}(A) = \{x \in X \mid f(x) \in A\} \in \mathcal{A}.$$

$\boxed{\text{ex}}$ $f(x) = 3.$ $\quad (X, \mathcal{A}) \xrightarrow{f} (\mathbb{R}, \mathcal{B}(I))$

Suffices to show for any $I \subseteq \mathbb{R}$ $I = (a, \infty)$.

$$f^{-1}(I) = \{x \in X \mid f(x) > a\} \in \mathcal{A}.$$

- For $a \geq 3$, $\{x \in X \mid 3 > a \geq 3\} = \emptyset \in \mathcal{A}.$
- For $a < 3$, $\{x \in X \mid a < 3\} = X \in \mathcal{A}.$

$\boxed{\text{Def}}$ <u>Random variable</u> $X: \Omega \to A$

is a measurable function from the sample space to a value in A.

$\boxed{A = \mathbb{R} \Rightarrow \text{continuous random variable}}$ $\quad$ $\boxed{\Omega \text{ discrete}, \ \mathcal{A} = 2^{\Omega} \Rightarrow \text{discrete random variable}}$

$\boxed{\text{example}}$ $\Omega = \{H, T\}$

$$Y(\omega) = \begin{cases} 1 & \text{if } \omega = H \\ 0 & \text{if } \omega = T \end{cases}$$

$$\mathbb{P}[X \in A] = \mathbb{P}[\{\omega \in \Omega \mid X(\omega) \in A\}]$$

$$= \mathbb{P}[\omega \in X^{-1}(A)] = \mathbb{P}[X^{-1}(A)]$$