2.1 Probability and measure

**Definition 2.1.** A measure $\mu$ on $(E, \mathcal{E})$ is a function $\mu : \mathcal{E} \rightarrow [0, \infty]$ such that if $(A_n, n \in \mathbb{N})$'s are disjoint then $\mu(\bigcup_{n \in \mathbb{N}} A_n) = \sum_{n \in \mathbb{N}} \mu(A_n)$. Then $(E, \mathcal{E}, \mu)$ is called a measure space. If $\mu(E) = 1$, then $\mu$ is called a probability measure and $(E, \mathcal{E}, \mu)$ is called a probability space.

**Example 2.2.** Let $E$ be finite, say $E = \{1, 2, \ldots, n\}$, and $\mathcal{E} = 2^E, \mu(A) := |A|, P(A) = |A|/n$ for $A \in \mathcal{E}$. $\mu$ is a measure on $(E, \mathcal{E})$ called the counting measure and $P$ is a probability measure on $(E, \mathcal{E})$ called the uniform probability measure.

A probability space $(E, \mathcal{E}, P)$ has outcomes $\omega \in E$ and events $A \in \mathcal{E}$.

**Theorem 2.3** (Properties of probability). From the definition of probability measure it follows that:

a. $P(\emptyset) = 0$

b. $P(A^c) = 1 - P(A)$

c. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

2.2 Random variables

**Definition 2.4.** Given a probability space $(\Omega, \mathcal{F}, P)$ and a measurable space $(E, \mathcal{E})$, an $(E, \mathcal{F})$-measurable random variable (rv) is a measurable function $X : \Omega \rightarrow E$, i.e., $X^{-1}(A) \in \mathcal{F}$ for all $A \in \mathcal{E}$.

**Special case:** If $E$ is countable, then $X : \Omega \rightarrow E$ is a $(2^E, \mathcal{F})$-measurable rv iff $X^{-1}(i) \in \mathcal{F}$ for all $i \in E$. The probability mass function (pmf) or distribution of $X$ is defined by

$$\lambda(i) := P(X = i) = P(\{\omega \in \Omega \mid X(\omega) = i\}).$$

**Example 2.5.** Flip 2 fair coins. $\Omega = \{HH, HT, TH, TT\}$. Define $E = \{0, 1, 2\}$ and $X :=$ number of heads in the 2 coin flips. Then $X$ is a $(2^E, 2^\Omega)$-measurable rv and

$$\lambda(0) = P(\{TT\}) = 1/4, \lambda(1) = P(\{HT, TH\}) = 1/2, \lambda(2) = P(\{HH\}) = 1/4.$$

Given a collection of $(\mathcal{E}, \mathcal{F})$-measurable random variable $X_i, i \in I$, we define $\sigma(X_i, i \in I)$ as the smallest $\sigma$-algebra w.r.t. which all $X_i, i \in I$'s are measurable, i.e.,

$$\sigma(X_i, i \in I) = \sigma(X_i^{-1}(A), i \in I, A \in \mathcal{E}).$$

In Example 2.5, $\sigma(X) = \sigma(\{TT\}, \{HH\}, \{HT, TH\})$. 
2.3 Conditional probability

**Definition 2.6.** If $A, B$ are events with $P(B) > 0$, the conditional probability of $A$ given $B$ is defined as

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

Note that $P(B \mid B) = 1$.

**Theorem 2.7 (Law of total probability).** If $B_n, n \in I$ is a partition of $E$ then for $A \in \mathcal{E}$ we have

$$P(A) = \sum_{i \in I} P(A \mid B_n) P(B_n).$$

**Definition 2.8.** Two events $A, B$ are independent iff $P(A \cap B) = P(A) P(B)$, equivalently $P(A \mid B) = P(A)$ when $P(B) > 0$.

Let $(E, \mathcal{E}, P)$ be a probability space. Assume that $E$ is countable. Thus there exists a partition $\Pi = \{E_i, i \in I\}$ generating $\mathcal{E}$.

**Definition 2.9.** Given a $\sigma$-algebra $\mathcal{B} \subseteq \mathcal{E}$ and an event $A \in \mathcal{E}$, the conditional probability of $A$ given $\mathcal{B}$ is a rv that is constant on each $E_i$ and

$$P(A \mid \mathcal{B})(\omega) := P(A \mid E_i) \text{ for all } \omega \in E_i.$$

**Example 2.10.** Toss 2 fair coins. Thus $E = \{HH, HT, TH, TT\}$ with $\mathcal{E} = 2^E, P(A) = \frac{|A|}{4}$. Let $\mathcal{B} = \{\emptyset, \{HH, HT\}, \{TH, TT\}, \Omega\}$. The partition $E_1 = \{HH, HT\}, E_2 = \{TH, TT\}$ generates $\mathcal{B}$.

Then

$$P(A \mid \mathcal{B}) = \begin{cases} P(A \mid \{HH, HT\}) & \text{if } \omega \in E_1, \text{ i.e., first coin toss is } H \\ P(A \mid \{TH, TT\}) & \text{if } \omega \in E_2, \text{ i.e., first coin toss is } T. \end{cases}$$

If $A = \{HH\}$, then $P(A \mid \mathcal{B}) = 1/2$ if first coin toss is $H$ and 0 otherwise. Note that, by law of total probability we have

$$P(A) = P(A \mid E_1) P(E_1) + P(A \mid E_2) P(E_2) = 1/2 \cdot 1/2 = 1/4.$$