1 Events and Probability

A **probability space** is a **measure space** $(\Omega, \mathcal{F}, \mathbf{P})$ consisting of:

- the **sample space** Ω — a set of outcomes called **sample**;
- the **σ-algebra** $\mathcal{F}$ — a family of subsets of Ω, called **events**, such that $\Omega \in \mathcal{F}$ and $\mathcal{F}$ is closed under complements (i.e. $\forall A \in \mathcal{F}$, $\Omega \backslash A \in \mathcal{F}$) and countable unions (i.e. $\forall A_i \in \mathcal{F}$, $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$);
- the **probability function** $\mathbf{P} : \mathcal{F} \to [0, 1]$ such that $\mathbf{P}(\Omega) = 1$ and $\mathbf{P}$ is **σ-additive** (i.e. $\mathbf{P}\left(\bigsqcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbf{P}(A_i)$).

The motivation behind this complicated definition is that some sets are **non-measurable**, thus mathematicians developed the theory of **measure**. For instance, **Borel set** on real line forms a σ-algebra which is **generated by** open intervals. **Stieltjes measure** is a **Borel measure** and builds the measure-theoretic foundation of **continuous probability distribution**.

Lemma 1.1 (Inclusion-exclusion principle) Let $E_1, \cdots, E_n$ be any n events. Then

$$\mathbf{P}\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{\ell=1}^{n} (-1)^{\ell+1} \sum_{i_1 < i_2 < \cdots < i_\ell} \mathbf{P}\left(\bigcap_{r=1}^{\ell} E_{i_r}\right).$$

Events $E_1, E_2, \cdots, E_n$ are **mutually independent** (simply called **independent** when $k = 2$) if and only if, for any subset $I \subseteq \{1, 2, \cdots, k\}$, $\mathbf{P}\left(\bigcap_{i \in I} E_i\right) = \prod_{i \in I} \mathbf{P}(E_i)$. Note that events $X, Y, Z, \cdots$ are unnecessarily mutually independent when they are pairwise independent.

The **conditional probability** that event E occurs given that event F occurs is $\mathbf{P}(E \mid F) = \mathbf{P}(E \cap F) \,/\, \mathbf{P}(F)$.

Theorem 1.2 (Law of total probability) Let events $\bigsqcup_{i=1}^{n} E_i = \Omega$. Then we have $\mathbf{P}(B) = \sum_{i=1}^{n} \mathbf{P}(B \mid E_i) \cdot \mathbf{P}(E_i)$.

Theorem 1.3 (Bayes's law) Let events $E_1, E_2, \cdots, E_n$ satisfy $\bigsqcup_{i=1}^{n} E_i = \Omega$. Then we have

$$\mathbf{P}(E_k \mid B) = \frac{\mathbf{P}(E_k \cap B)}{\mathbf{P}(B)} = \frac{\mathbf{P}(B \mid E_k) \cdot \mathbf{P}(E_k)}{\sum_{i=1}^{n} \mathbf{P}(B \mid E_i) \cdot \mathbf{P}(E_i)}.$$

In the **Bayesian approach** one starts with a **prior** model, giving some initial value to the model parameters. This model is then modified, by incorporating new observations, to obtain a **posterior** model that captures the new information.

Exercise 1.6 Using mathematical induction, we have $p_{i,j} = \frac{i-1}{i+j-1} \cdot p_{i-1,j} + \frac{j-1}{i+j-1} \cdot p_{i,j-1} = \frac{i+j-2}{i+j-1} \cdot \frac{1}{i+j-2} = \frac{1}{i+j-1}$.

Exercise 1.7.b Let $F_{b_1 b_2 \cdots b_n}$ be the intersection of events $E_i\,(b_i = 1)$ or $\Omega \backslash E_i\,(b_i = 0)$, and P_k be the sum of $\mathbf{P}(F_b)$ where b consists of k one and $n - k$ zero. Then for every $k \geq 1$, we have $\sum_{i=1}^{l} (-1)^{i+1} \binom{k}{i} = 1 + (-1)^{l+1} \binom{k-1}{l} \geq 1$. Multiply both sides by P_k and sum them up. We eventually reach the desired inequality.

Exercise 1.11.b $p_3 = p_1 \cdot (1 - p_2) + (1 - p_1) \cdot p_2 \Rightarrow q_3 = 1 - 2p_3 = (1 - 2p_1)(1 - 2p_2) = q_1 q_2$. Is there any underlying motivation?

Exercise 1.24 (Karger's algorithm) Let K be the minimum r-way cut-set. Considering all r-way cut-sets consisting of $r - 1$ single vertex, the total size is $m \cdot \binom{n-2}{r-1}$ with an upper bound $(m - |K|) \cdot \binom{n}{r-1}$. It follows that

$$m \cdot \binom{n-2}{r-1} \leq (m - |K|) \cdot \binom{n}{r-1} \quad \Rightarrow \quad 1 - \frac{|K|}{m} \geq \binom{n-2}{r-1} \binom{n}{r-1}^{-1} = \frac{(n-r+1)(n-r)}{n(n-1)}.$$

The probability that K survives all the $n - r$ iterations is at least

$$\prod_{i=0}^{n-r-1} \frac{(n-i+1-r)(n-i-r)}{(n-i)(n-i-1)} = r \cdot \binom{n}{r-1}^{-1} \binom{n-1}{r-1}^{-1}$$

and its reciprocal is the maximum possible number of minimum cardinality of r-way cut-sets.