

1. A **probability space** is a measurable space (S, \mathcal{M}, μ) with $\mathcal{M} \subseteq \mathcal{P}(S)$ and $\mu: \mathcal{M} \rightarrow [0, 1]$ with $\mu(S) = 1$. In particular, we will write $\mathcal{M} = \mathcal{F}$, $\Omega = S$ and $\mu = \mathbb{P}$. The sets $E \in \mathcal{M}$ are called **events**. The σ -algebra \mathcal{F} must satisfy the following axioms:

- (a) $\emptyset \in \mathcal{F}$
- (b) If $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$
- (c) If $A_1, A_2, \dots \in \mathcal{F}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

As well, \mathbb{P} must satisfy

- (a) $\mathbb{P}(A) \geq 0$ for all $A \in \mathcal{F}$.
- (b) $\mathbb{P}(\Omega) = 1$.
- (c) Let A_1, A_2, \dots be disjoint. Then $\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.

As a result, we collect some general properties here:

- (a) $\mathbb{P}(A^c) + \mathbb{P}(A) = 1$
- (b) $\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{r=1}^n (-1)^{r+1} \sum_{1 \leq i_1 < \dots < i_r \leq n} \mathbb{P}(A_{i_1} \cap \dots \cap A_{i_r})$
- (c) We say D_1, D_2, \dots is a **decreasing** (resp. **increasing**) sequence of events if $D_1 \supset D_2 \supset \dots$ (resp. $D_1 \subset D_2 \subset \dots$) and define $\lim_{n \rightarrow \infty} D_n = \bigcap_{i=1}^{\infty} D_n$ (resp. $\lim_{n \rightarrow \infty} D_n = \bigcup_{i=1}^{\infty} D_n$). Then $\mathbb{P}(\lim_{n \rightarrow \infty} D_n) = \lim_{n \rightarrow \infty} \mathbb{P}(D_n)$.

2. Let $E, F \in \mathcal{F}$. Fix E with $\mathbb{P}(E) > 0$. Then define $\mathbb{P}(F|E) = \frac{\mathbb{P}(F \cap E)}{\mathbb{P}(E)}$, the **conditional probability** of F given E . In fact, the map $\mathbb{P}(\cdot|E): \mathcal{F} \rightarrow \mathbb{R}$ is a probability measure in its own right. Conditional probability has some nice properties:

- (a) $\mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_n) = \mathbb{P}(A_1|A_2 \cap \dots \cap A_n) \cdots \mathbb{P}(A_3|A_2 \cap A_1) \mathbb{P}(A_2|A_1) \mathbb{P}(A_1)$.
- (b) Let $A_1, A_2, \dots, A_n \in \mathcal{F}$ be a partition of Ω . Then for any $B \in \mathcal{F}$,

$$\mathbb{P}(B) = \sum_{i=1}^n \mathbb{P}(B|A_i) \mathbb{P}(A_i)$$

and

$$\mathbb{P}(A_k|B) = \frac{\mathbb{P}(B|A_k) \mathbb{P}(A_k)}{\sum_{i=1}^n \mathbb{P}(B|A_i) \mathbb{P}(A_i)}$$

3. Let $A_1, A_2, \dots, A_n \in \mathcal{F}$. We say the events are **independent (as a collection)** if

$$\mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \cdots \mathbb{P}(A_{i_k})$$

We say the events are **conditionally independent** with respect to B if they are independent with respect to the conditional probability measure:

$$\mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k}|B) = \mathbb{P}(A_{i_1}|B) \cdots \mathbb{P}(A_{i_k}|B)$$

4. A **random variable** is a measurable function $f : S \rightarrow \mathbb{R}$. We usually write random variables with capital letters, like X . We say X is **discrete** if its range is countable.

Let \mathcal{B} denote the Borel measurable subset of \mathbb{R} . The **distribution** of a measurable function is a measure $\rho : \mathcal{B} \rightarrow \mathbb{R}$ given by $\rho(S) = \mathbb{P}(f^{-1}(S))$. This is indeed measurable since: by definition, X is measurable if it pulls open sets in \mathbb{R} to elements of \mathcal{F} .

For some reason, probabilists dislike writing sets down properly. We say $\mathbb{P}(X = i) = \mathbb{P}(X^{-1}(i))$, $\mathbb{P}(X \leq i) = \mathbb{P}(X^{-1}(-\infty, i])$, and other similar notations.

The **expected value** of a random variable X is given by $(X) = \int_{\Omega} X d\mathbb{P}$. When X is discrete so $X(\Omega) = \{x_1, x_2, \dots\}$, this becomes $(X) = \sum_{k=1}^{\infty} x_k \mathbb{P}(X = x_k)$.

(a) If X is discrete and $g : \mathbb{R} \rightarrow \mathbb{R}$, then $(g(X)) = \sum_{k=1}^{\infty} g(x_k) \mathbb{P}(X = x_k)$.

(b) $(aX + Y) = a(X) + (Y)$.

We define the **variance** $(X) = [(X - \mu)^2]$.

(a) $(X) = (X^2) - (X)^2$