Probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a special kind of general measure space. We require

$$\mathbb{P}(\Omega) = 1.$$

The finiteness of probability measure brings some convenient properties.

1. *the continuity of probability measure.*
   If $A_k \uparrow A$, then $\mathbb{P}(A_k) \uparrow \mathbb{P}(A)$.
   If $A_k \downarrow A$, then $\mathbb{P}(A_k) \downarrow \mathbb{P}(A)$.
2. *there are at most countable atoms.*
$$D \coloneqq \left\{\omega \in \Omega \,\middle|\, \mathbb{P}\{\omega\} > 0\right\} \quad \text{is at most countable}.$$
3. *boundness implies integrability.*
$$|X| \leq M < \infty \quad \Rightarrow \quad \mathbb{E}\,|X| < \infty.$$
4. *high order moment exists* $\quad \Rightarrow \quad$ *low order moment exists.*
$$\mathbb{E}\,|X|^q < \infty \quad \Rightarrow \quad \mathbb{E}\,|X|^p < \infty, \qquad \forall 1 \leq p \leq q < \infty.$$

It is worth noting that we often require $(\Omega, \mathcal{F}, \mathbb{P})$ is **complete**, i.e.,

$$\mathbb{P}(N) = 0 \quad \Rightarrow \quad \forall N_0 \subset N, N_0 \in \mathcal{F}.$$

A **random variable (r.v.)** X refers to a measurable function from $(\Omega, \mathcal{F})$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

The **distribution** of X refers to the probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ induced by it,

$$\mu_X(B) := \mathbb{P}[X \in B].$$

Statistics of X: with μ_X, we can easily predict probabilistic behaviors of X,

$$\mathbb{E}\, X = \int_{\mathbb{R}} x\, \mu_X(dx),$$
$$\mathbb{E}\, X^2 = \int_{\mathbb{R}} x^2\, \mu_X(dx).$$

Generally, for a Borel function $g : \mathbb{R} \to \mathbb{R}$,

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}} g(x)\, \mu_X(dx).$$

This equality should be understood that

1. the existence of one integral implies the existence of the other integral;
2. if these integral exist, then they must be equal.

The **(cumulative) distribution function** of a r.v. X refers to a nondecreasing function $F_X : \mathbb{R} \to [0,1]$,

$$F_x(x) := \mu_X(-\infty, x].$$

Note that F_X is **right continuous** in our definition. It could be **left continous** if F_X is defined by $\mu_X(-\infty, x)$.

A r.v. X is called **discrete** if there exists a countable set $D \subset \mathbb{R}$ such that $\mu_X(D) = 1$.
Equivalently, this means X has at most countable possible values.

A r.v. X is called **(absolutely) continuous** if $\mu_X \ll \lambda$.
Equivalently, this means (by Radon-Nikodym theorem) there exists $f_X \geq 0$ such that

$$\mu_X(B) = \int_B f_X(x)\, dx, \qquad \forall B \in \mathcal{B}(\mathbb{R}).$$

In this case, f_X, also known as the R-N derivative, is called the **probability density function (p.d.f.)** of X.

Statistics of continuous r.v. X: for a Borel function $g : \mathbb{R} \to \mathbb{R}$,

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}} g(x) f_X(x)\, dx.$$

Note: if X is generalized as a measurable function from $(\Omega, \mathcal{F})$ to $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, then

1. μ_X is a probability measure on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$;
2. F_X and f_X (if exists) are functions on $\mathbb{R}^n$.

Measurable set classes $\{\mathcal{A}_t\}_{t\in T}$ are termed **independent** if for any *finite* subset $\{t_1, t_2, \dots, t_n\} \subset T$,

$$\mathbb{P}\Big(\bigcap_{k=1}^{n} A_{t_k}\Big) = \prod_{k=1}^{n} \mathbb{P}(A_{t_k}), \qquad \forall A_{t_k} \in \mathcal{A}_{t_k}, \quad k = 1, 2, \dots, n.$$

This definition naturally applies to r.v.s $\{X_t\}_{t\in T}$, where $\mathcal{A}_t$ is replaced by $\sigma(X_t) \coloneqq \{X_t^{-1}(B) \,|\, B \in \mathcal{B}(\mathbb{R})\}$.

Specifically, *finite* many r.v.s $\{X_1, X_2, \dots, X_n\}$ are independent if and only if

$$\mathbb{P}\Big(\bigcap_{k=1}^{n} [X_k \le x_k]\Big) = \prod_{k=1}^{n} \mathbb{P}[X_k \le x_k], \qquad \forall x_k \in \mathbb{R}, \quad k = 1, 2, \dots, n.$$

Independent-means-multiply: given independent variables X and Y, if $X, Y \in \mathcal{L}^1$ or $X, Y \in \mathcal{L}^0_+$, then

$$\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y].$$

Counterexample: it is possible that XY is not well defined when X, Y are independent and $X \in \mathcal{L}^1, Y \in \mathcal{L}^0_+$. We can construct such an example following this procedure:

1. fix $\Omega = [0,1]$ and $\mathbb{P}(B) = \lambda(B \cap [0,1])$;
2. set $X(\omega) = \mathbb{1}_{[0,1/2]}(\omega) - \mathbb{1}_{(1/2,1]}(\omega) \in \mathcal{L}^1$;
3. set $Y \in \mathcal{L}^0_+$ such that $\mathbb{E}\,Y = \infty$;
4. show that X and Y are independent if $\forall \omega \in [0,1],\ Y(\omega) = Y(1-\omega)$;
5. the last step uses the fact that Lebesgue measure is invariant under reflection and translation.

A sequence of probability measures $\{\mu_n\}$ on $\mathcal{B}(\mathbb{R})$ is said to **weakly converge** to a probability measure μ if

$$\int_{\mathbb{R}} f(x)\,\mu_n(dx) \to \int_{\mathbb{R}} f(x)\,\mu(dx), \qquad \text{for any bounded continous function } f : \mathbb{R} \to \mathbb{R}.$$

A sequence of r.v.s $\{X_n\}$ is said to **converge to X in distribution (or in law)** if

$$\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)], \qquad \text{for any bounded continous function } f : \mathbb{R} \to \mathbb{R}.$$

Note: X_n convenges to X in distribution, denoted by $X_n \xrightarrow{\mathcal{D}} X$ or $X_n \xrightarrow{d} X$,
is equivalent to μ_{X_n} weakly converges to μ_X, denoted by $\mu_{X_n} \xrightarrow{w} \mu_X$.

The **characteristic function of a probability measure** μ on $\mathcal{B}(\mathbb{R})$ is a complex-valued function defined on $\mathbb{R}$

$$\phi_\mu(t) := \int_{\mathbb{R}} e^{itx}\,\mu(dx) := \int_{\mathbb{R}} \cos tx\,\mu(dx) + i \int_{\mathbb{R}} \sin tx\,\mu(dx).$$

The **characteristic function of a r.v.** X on $\mathcal{B}(\mathbb{R})$ is a complex-valued function defined on $\mathbb{R}$

$$\phi_X(t) := \mathbb{E}[e^{itX}] := \mathbb{E}[\cos tX] + i\,\mathbb{E}[\sin tX].$$

We should note that the **characteristic function is bounded by 1**.

We should also bear in mind that **convergence in law is equivalent to pointwise convergence of characteristic function**.

(The implication in one direction is by definition, and the opposite one comes from Lévy's continuity theorem.)