Metric-Measure Spaces

In this brief overview, we present the basics of measure theory with particular emphasis on metric spaces endowed with measure structure.

**Sigma algebra.** Let $X$ be a set. The power set of $X$ is the set $2^X = \{A : A \subset X\}$ of all subsets of $X$. A subset $\Sigma \subseteq 2^X$ is called a \(\sigma\)-algebra over $X$ if it satisfies the following properties:

1. *Non-emptiness:* $\Sigma \neq \emptyset$.
2. *Closure under complement:* for every $A \in \Sigma$, $A^c = X \setminus A \in \Sigma$.
3. *Closure under countable union:* for every countable collection $\{A_i : A_i \in \Sigma\}$, $\bigcup A_i \in \Sigma$.

**Measure.** Let $X$ be a space, and $\Sigma$ a \(\sigma\)-algebra over it. A function $\mu : \Sigma \to \mathbb{R} \cup \{\infty\}$ is called a measure over $\Sigma$ if it satisfies the following properties:

1. *Non-negativity:* $\mu(A) \geq 0$ for every $A \in \Sigma$.
2. *Null empty set:* $\mu(\emptyset) = 0$.
3. *Countable additivity:* for every countable collection $\{A_i : A_i \in \Sigma\}$ of pairwise-disjoint sets (i.e., $A_i \cap A_j = \emptyset$ for every $i \neq j$),

$$\mu\left(\bigcup A_i\right) = \sum \mu(A_i).$$

A set $A \in \Sigma$ is called a measurable set, and the triplet $(X, \Sigma, \mu)$ is called a measure space. For brevity, a measure space is often denoted as $(X, \mu)$ or simply $X$, implying the existence of some \(\sigma\)-algebra defining measurable sets.

A certain point-wise property on $X$ is said to be satisfied almost everywhere (abbreviated as “a.e.”) if it is satisfied on $A \subset X$ with $\mu(A) = \mu(X)$ (or, in other words, it is violated on a set with measure zero).

A measure satisfies the following properties:

1. *Monotonicity:* $\mu(A_1) \leq \mu(A_2)$ for every measurable sets $A_1 \subseteq A_2$.
2. *Countable sub-additivity:* for every countable collection $\{A_i\}$ of measurable sets,

$$\mu\left(\bigcup A_i\right) \leq \sum \mu(A_i)$$

(the inequality holds with equality if the sets are pairwise-disjoint or the measure of $A_i \cup A_j$ is zero for every $i \neq j$).
3. **Continuity**: For a measurable set \( A_1 \subseteq A_2 \subseteq \cdots \)
\[
\mu \left( \bigcup A_i \right) \leq \sum \mu(A_i)
\]

**Probability space.** A measure is called finite if \( \mu(X) < \infty \). If, furthermore, \( \mu(X) = 1 \), \( \mu \) is called a probability measure. A measure space \( (X, \Sigma, \mu) \) with a probability measure is usually called a probability space. In such cases, the set \( X \) is referred to as the sample space representing outcomes; the \( \sigma \)-algebra \( \Sigma \) is referred to as the set of events where each event contains zero or more outcomes; and the probability measure \( \mu \) assigns the probability of each event.

**Lebesgue integral.** A measure is a formal construction assigning to each measurable set a number, which can be interpreted as the set size. This generalizes the notions of length, area, and volume. In other words, measures are instrumental in the formalization of the notion of an integral.

Let \( (X, \Sigma, \mu) \) be a measure space. A real function \( f : X \to \mathbb{R} \) is said to be measurable if the preimage of every closed interval is a measurable set, i.e., \( f^{-1}([a, b]) = \{ x \in X : f(x) \in [a, b] \} \in \Sigma \). The Lebesgue integral (sometimes simply the integral) of a measurable real function \( f \) w.r.t. \( \mu \) is denoted by
\[
\int f, \int fd\mu, \int_X f(x)d\mu(x), \text{ or } \int_X f(x)\mu(dx)
\]
and constructed as follows:

**Indicator functions**: the integral of an indicator function \( 1_A \) associated with a measurable set \( A \) is given simply by
\[
\int 1_A d\mu = \mu(A).
\]

**Simple functions**: the integral of a linear combination \( s = \sum a_k 1_{A_k} \) of indicator functions (a.k.a. a simple function) is given by
\[
\int \left( \sum_k a_k 1_{A_k} \right) d\mu = \sum_k a_k \int 1_{A_k} d\mu = \sum_k a_k \mu(A_k).
\]

**Non-negative functions**: the integral of a non-negative measurable function \( f \) is given by
\[
\int fd\mu = \sup \left\{ \int sd\mu : 0 \leq s \leq f, s \text{ simple} \right\}.
\]

**Arbitrary measurable functions**: finally, given an arbitrary measurable function \( f \), it can be decomposed into non-negative measurable functions
\[
f^+(x) = \begin{cases} f(x) & : f(x) > 0 \\ 0 & : \text{else} \end{cases}
\]

\[2\]
and

\[ f^-(x) = \begin{cases} -f(x) & : f(x) < 0 \\ 0 & : \text{else.} \end{cases} \]

In this decomposition, \( f = f^+ - f^- \) and \( |f| = f^+ + f^- \). If \( |f| \) has a finite integral, then \( f \) is said to be integrable and its integral is defined as

\[ \int f \, d\mu = \int f^+ \, d\mu - \int f^- \, d\mu. \]

The Lebesgue integral satisfies the following properties:

1. **Indistinguishability:** if \( f = g \) a.e., then

\[ \int f \, d\mu = \int g \, d\mu. \]

2. **Linearity:** for any real \( a \) and \( b \),

\[ \int (af + bg) \, d\mu = a \int f \, d\mu + b \int g \, d\mu. \]

3. **Monotonicity:** if \( f \leq g \), then

\[ \int f \, d\mu \leq \int g \, d\mu. \]

4. **Monotone convergence:** let \( \{f_k\} \) be a sequence of non-negative measurable functions such that \( f_k \leq f_{k+1} \) for every \( k \). Then,

\[ \lim_{k \to \infty} \int f_k \, d\mu = \int \lim_{k \to \infty} f_k \, d\mu. \]

**Measure coupling.** Let \((X, \Sigma, \mu)\) and \((Y, \Omega, \nu)\) be two measure spaces. The product space \( X \times Y \) can be trivially equipped with the product \( \sigma \)-algebra \( \Sigma \times \Omega \), on which a measure \( \rho \) can be defined. The **marginal** of \( \rho \) on the first factor is defined as

\[ \rho(A|Y) = \int_{X \times Y} 1_A \, d\rho \]

for every \( A \in \Sigma \). In the same way, the marginal \( \rho(B|X) \) on the second factor is defined. A measure \( \rho \) on \( X \times Y \) is said to be a coupling of \( \mu \) and \( \nu \) if \( \rho(A|Y) = \mu(A) \) and \( \rho(B|X) = \nu(B) \) for every measurable sets \( A \in \Sigma \) and \( B \in \Omega \). A particular example of a coupling of \( \mu \) and \( \nu \) is the **product measure** \( \rho = \mu \times \nu \) defines as the (unique) measure satisfying \( (\mu \times \nu)(A \times B) = \mu(A)\nu(B) \) for every \( A \in \Sigma \) and \( B \in \Omega \).
Borel space. Recall that a topology $\mathcal{T}$ on $X$ is a subset of the power set $2^X$ that is closed under union and finite intersection and contains the empty set and $X$ itself. Elements of $\mathcal{T}$ are called open sets and the pair $(X, \mathcal{T})$ is called a topological space. The smallest $\sigma$-algebra on a topological space containing all open sets is called the Borel $\sigma$-algebra and is denoted by $\mathcal{B}(X) \supseteq \mathcal{T}$. A topological space equipped with the Borel $\sigma$-algebra is called a Borel space. Measures define on such a space are called Borel measures. This is a “canonical” way of equipping a topological space with a measure structure.

Metric-measure space. Recall that a metric space is automatically a topological space since a metric induces a topology (where open sets are generated by open metric balls). This allows to equip a metric space with the Borel $\sigma$-algebra, on which measures can be defined. A metric space $(X, d)$ equipped with a Borel measure $\mu$ is called a metric-measure or mm space. If $\mu$ is further a probability measure, the quantity

$$m_p(x_0) = \int d(x, x_0)^p d\mu(x)$$

is called the $p$-th moment of $\mu$ about the point $x_0 \in X$. If $X$ is compact, all $p$-th moments are finite and bounded above by $\text{diam}(X)^p$.

Wasserstein distance. Let $(X, d)$ be a compact metric space equipped with a Borel $\sigma$-algebra and let us fix $p \geq 1$. Then, the Wasserstein distance between two probability measures $\mu$ and $\nu$ on $X$ is defined as

$$D_p(\mu, \nu) = \left( \inf_\rho \int_{X \times X} d(x, x')^p d\rho(x, x') \right)^{1/p},$$

where the infimum is taken over all couplings $\rho$ of $\mu$ and $\nu$. This defines a family of metrics on the space of probability measures on $X$.

Intuitively, the two measures $\mu$ and $\nu$ can be visualized as, respectively, a pile of earth and a hole in the ground, where “$\mu(x)$” is the amount of earth located over $x$, and “$\nu(x')$” is the capacity of the hole below a point $x'$. A coupling $\rho$ can be thought of a transportation plan: it tells to transfer “$\rho(x, x')$” units of earth from point $x$ to point $x'$. The cost of $d(x, x')^p$ is assigned to transferring a unit of earth from $x$ to $x'$. Our goal is to transport all earth from the pile to the hole minimizing the overall transportation cost. Wasserstein distances essentially solve this problem by finding the minimum-cost transportation plan. Because of this interpretation, Wasserstein distances are widely known as earth mover’s distances or EMD for short.