# Measure Theory

Part III

## Contents

## 3.1 Probability Space

We recall the definition of a probability space as the triple:

$$(\Omega,\ \mathscr{F},\ P)$$

$\Omega$ — **Sample Space.** The set of all possible outcomes of a system or experiment. For a fair die:

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

$\mathscr{F}$ — **Event Space.** A collection of events, where events are subsets of outcomes, $\mathscr{F} \subseteq \mathcal{P}(\Omega)$, to which a probability is assigned. An outcome may belong to different events, and events may contain different subsets of outcomes.

The event space is structured as a **sigma-algebra**. When we define an event of interest, we implicitly select the minimal sigma-algebra such that the rules of probability hold. Defining an event means choosing what that event is, what it is not, what is trivial ($\Omega$), and what is null ($\varnothing$). Numerous sigma-algebras can exist, corresponding to many definitions of possible events. The choice of event is implicit with the choice of $\mathscr{F}$.

**Note.** We also define countable union (and its complement, countable intersection) to make the combination of subsets and events meaningful probabilistically.

**Examples of events for a die:**

- Rolling low numbers: $\{1, 2, 3\}$
- Rolling evens: $\{2, 4, 6\}$
- Rolling a 2: $\{2\}$

For each event, a sigma-algebra can be chosen that permits it while respecting the axioms.

If we define *even* as $\{2, 4, 6\}$, we must also include its complement *odd* $= \{1, 3, 5\}$, the empty event $\varnothing$, and the full event $\Omega$.

If we define the event as rolling a 2, subset $\{2\}$, we must also include its complement $\{1, 3, 4, 5, 6\}$.

$P$ — **Probability Measure.** A measure function that assigns a "size" to elements of $\mathscr{F}$. It obeys Kolmogorov's axioms and is used to define probability.

## 3.2 Borel Algebra

### 3.2.1 Topology

A **topology** $\mathscr{T}$ on a set $X$ is a subset of $\mathcal{P}(X)$ such that the elements of $X$ are open. This requires both $\varnothing$ and $X$ to be in $\mathscr{T}$, and for $\mathscr{T}$ to be closed under arbitrary union and finite intersection. The elements of $\mathscr{T}$ are called **open sets**.

Open sets allow for the definition of nearness of points relative to neighbours without invoking the notion of distance via a metric.

### 3.2.2 $\sigma$-Algebra

Let $X$ be a set and $\mathcal{P}(X)$ its powerset. A subset $\Sigma \subseteq \mathcal{P}(X)$ is a **sigma-algebra** on $X$ if it satisfies:

1. $X \in \Sigma$
2. Closed under complementation: $A \in \Sigma \Rightarrow A^c \in \Sigma$

3. Closed under countable union: $A_1, A_2, \ldots \in \Sigma \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \Sigma$

From these axioms it follows that $\varnothing \in \Sigma$, and by De Morgan's law, $\Sigma$ is also closed under countable intersection.

### 3.2.3 Borel $\sigma$-Algebra

Previously, sigma-algebras were defined on regular sets $X$. We now extend by defining sigma-algebras on a topology $\mathscr{T}$. Since $\mathscr{T} \subseteq \mathcal{P}(X)$, we can form a sigma-algebra from the members of $\mathscr{T}$. This is the **Borel sigma-algebra** $\mathscr{B}(\mathscr{T})$ or $\mathscr{B}(X)$, whose elements are called **Borel sets** $B$.

Through the application of sigma-algebra axioms, $\mathscr{B}(\mathscr{T})$ can contain elements of $X$ not present in the topology $\mathscr{T}$ itself.

**Example.** Let $X = \mathbb{R}$ with the standard topology $\tau$ (open intervals and their unions).

Single points such as $\{1\}$ are not in $\tau$, since no point is an open set in $\mathbb{R}$. However, consider the sequence of open sets:

$$U_n = \left(1 - \frac{1}{n},\ 1 + \frac{1}{n}\right) \in \tau \quad \forall n$$

By countable intersection:

$$\bigcap_{n=1}^{\infty} U_n = \bigcap_{n=1}^{\infty} \left(1 - \frac{1}{n},\ 1 + \frac{1}{n}\right) = \{1\}$$

So $\{1\} \in \mathscr{B}(\tau)$ via only the sigma-algebra axioms applied to $\tau$.

This is why the Borel sigma-algebra is the **smallest sigma-algebra containing** $\tau$. Larger sigma-algebras include the full powerset $\mathcal{P}(X)$ and the Lebesgue sigma-algebra $\mathscr{L}(X)$.

**Note.** Even though the topology contains only open sets, the derived requirement of closure under complements means $\mathscr{B}(\mathscr{T})$ also contains closed sets.

With an underlying set $X$, sigma-algebra $\mathscr{B}(\mathscr{T})$, and a measure $\mu$ (the **Borel measure**), we obtain a measure space. Taking $\mathbb{R}$ as the underlying set with the standard topology:

$$(\mathbb{R},\ \mathscr{B}(\mathscr{T}),\ \mu)$$

This is a full measure space. Defining measure spaces on $\mathbb{R}$ opens up the tools of real analysis.

## 3.3 Random Variable $X$

To apply real analysis, it is useful to transfer the measure defined on the probability space $(\Omega, \mathscr{F})$ to the real measurable space $(\mathbb{R}, \mathscr{B}(\mathscr{T}))$. This is accomplished through a **measurable function** — a **random variable** $X$:

$$X : \Omega \to \mathbb{R}$$

The measurability condition requires:

$$X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathscr{F}$$

This means that whatever the measurable function transfers onto the real line must live in the sigma-algebra $\mathscr{F}$. The **pushforward measure** $P_X$ induced by $X$ is:

$$P_X(B) = P\big(X^{-1}(B)\big)\,, \quad B \in \mathscr{B}(\mathbb{R})$$

The transfer of structure is:

$$(\Omega, \mathscr{F}, P) \xrightarrow{X} (\mathbb{R}, \mathscr{B}(\mathbb{R}), P_X)$$

$P_X$ is the measure from the probability space transferred to work on $\mathbb{R}$. The random variable also encodes structure preservation: $\mathscr{F}$ is ensured to be equivalent to $\mathscr{B}(\mathbb{R})$ under $X$.

**Remark.** When defining a random variable, the numeric choice is arbitrary — only structure is transferred. Setting the origin to 1000 instead of 0 leaves all other measurements relative to 1000 unchanged.

Random variables allow us to probe the same system differently. If $X$ is concerned with pressure, we can define $Y$ concerned with temperature, generating a different distribution. However, $Y$ requires its own sigma-algebra, since the sigma-algebras must match their respective domains according to the pushforward axiom.

## 3.4 CDF

Several functions can model $P_X$ when applying real analysis. The minimal faithful representation is the **Cumulative Distribution Function (CDF)**.

The real line $\mathbb{R}$ can be refined by building additional structures to support analysis.

#### *Order Theory*

The order structure on $\mathbb{R}$ defines intervals:

$$a < x < b \quad \iff \quad (a, b)$$

#### *Topology*

An open interval is a subset of $\mathbb{R}$ with no endpoints and no internal gaps — a fixed segment of the real line where points and their neighbours always remain inside. The standard topology on $\mathbb{R}$ is the collection of all open sets, generated by arbitrary unions of open intervals.

#### *Measure Theory*

The Borel sigma-algebra is generated from the open sets in $\mathscr{T}_{\mathbb{R}}$. Since open sets can be generated from open intervals, the Borel sigma-algebra originates from open intervals. Thus, defining $P_X$ on intervals is sufficient to define the measure on all of $\mathscr{B}(\mathbb{R})$.

Instead of two parameters $a, b$ to define intervals $(a, b]$, we use **half-lines**:

$$(a, b] = (-\infty, b] \setminus (-\infty, a]$$

Differences of half-lines recover intervals. Complements are:

$$(a, b]^c = (-\infty, a] \cup (b, \infty)$$

This gives a **one-parameter** method to describe intervals via half-lines $(-\infty, x]$.

We therefore define the CDF:

$$F(x) = P_X((-\infty, x])$$

**Out of scope, but notable:**

- The $\pi$-$\lambda$ theorem guarantees the *uniqueness* of $F(x)$ in determining $P_X$.

- Carathéodory's Extension Theorem guarantees the *existence* of a measure on $\mathscr{B}(\mathbb{R})$ consistent with the defined CDF.

## 3.5 PDF

When working on $\mathbb{R}$, we have two measures defined on the same measurable space $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$: the pushforward $P_X$ and the **Lebesgue measure** $\lambda$. We may wish to express one measure as a weighted version of the other — to "rescale" $P_X$ with respect to the natural measure of $\mathbb{R}$.

The **Radon-Nikodym Theorem** states: if $P_X$ is absolutely continuous with respect to $\lambda$, i.e. $P_X \ll \lambda$, then there exists a unique function $f$ such that:

$$P_X(B) = \int_B f \, d\lambda \quad \forall \, B \in \mathscr{B}(\mathbb{R})$$

$f$ is the **Radon-Nikodym derivative**:

$$f = \frac{dP_X}{d\lambda}$$

This function $f$ is the **PDF**.

**Relation to the CDF.** Since $\lambda((a, b)) = b - a$ and $dt$ denotes integration with respect to $\lambda$:

$$F(x) = P_X((-\infty, x]) = \int_{-\infty}^{x} f(t) \, d\lambda(t) = \int_{-\infty}^{x} f(t) \, dt$$

Therefore:

$$\frac{dF}{dx} = f(x)$$

The PDF is the derivative of the CDF.

## 3.6 Measures of Central Tendency & Dispersion

With the Lebesgue measure and integral established, expectation and variance follow naturally:

$$E[X] = \int_{\mathbb{R}} x \, dP_X(x)$$

$$\mathrm{Var}(X) = \int_{\mathbb{R}} (x - E[X])^2 \, dP_X(x)$$

## 4. Random Vectors

Let $(\Omega, \mathscr{F}, P)$ be a probability space. A **random vector** in $\mathbb{R}^n$ is a measurable map:

$$X : \Omega \to \mathbb{R}^n$$

with respect to $\mathscr{F}$ and the Borel $\sigma$-algebra on $\mathbb{R}^n$.

Equivalently, we have component random variables:

$$X_i : (\Omega, \mathscr{F}) \to (\mathbb{R}, \mathscr{B}(\mathbb{R})), \quad i = 1, \ldots, n$$

all defined on the same measurable space $(\Omega, \mathscr{F})$, and we form:

$$X(\omega) = \big(X_1(\omega), \, \ldots, \, X_n(\omega)\big)$$

**Key requirement.** For a valid random vector, all components must share the same domain $\Omega$ and the same $\sigma$-algebra $\mathscr{F}$. Otherwise, the tuple $(X_1, \ldots, X_n)$ is not a well-defined function on a single space.

### Different Domains

Suppose the component random variables are defined on different spaces:

$$X_1 : (\Omega_1, \mathscr{F}_1) \to \mathbb{R}, \qquad X_2 : (\Omega_2, \mathscr{F}_2) \to \mathbb{R}$$

We cannot directly form $(X_1, X_2)$. To combine them, we first move to a common space — the **product space**:

$$(\Omega_1 \times \Omega_2,\ \mathscr{F}_1 \otimes \mathscr{F}_2)$$

Then define:

$$X(\omega_1, \omega_2) = \big(X_1(\omega_1),\ X_2(\omega_2)\big)$$

which is a valid random vector on the product space.

### Summary

| **Case** | **Procedure** |
|---|---|
| Same domain $(\Omega, \mathscr{F})$ | Directly forms a random vector |
| Different domains | Embed into a common product space first |