Measure Theory (Part III)

3.1 Probability Space

We remember defining probability spaces

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

Ω: The sample space. These are all the possible outcomes of a system or results of an experiment. Example for a game of dice, $\Omega = \{1, 2, 3, 4, 5, 6\}$

$\mathcal{F}$: The event space. It contains events which are subsets of outcomes, $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ to which a probability is assigned. An outcome may be an element of different group of events, and events may include different group of outcomes.

The event space is structured as a sigma algebra. When we define an event we are interested in, we implicitly select for such a minimal sigma algebra such that the rules of probability holds. When we define an event we implicitly choose what that event is, and what it is not. What is everything/trivial, Ω, and what is null, $\emptyset$. Numerous instances of sigma algebras can exist, matching many definitions of possible events. The choice of event is implicit with the choice of $\mathcal{F}$.

Note: We also define countable union [and its complement; countable intersection] to make combination of subsets/events meaningful probabilistically.

An event could be: Rolling low numbers $\{1, 2, 3\}$, Rolling evens $\{2, 4, 6\}$, Rolling a 2 $\{2\}$.

For each of these events, we can choose a sigma algebra that allows for it, while respecting the axioms.

If we define even [forms a subset of $\mathcal{P}(\Omega)$] being $\{2, 4, 6\}$, we must define its complementary; odd $\{1, 3, 5\}$. We must also define rolling nothing, $\emptyset$, and rolling something, Ω.

If we define event as rolling a 2, subset of $\mathcal{P}(\Omega)$ being $\{2\}$, we must also define not 2, subset being $\{1, 3, 4, 5, 6\}$.

P: A measure function that assign "size" to the elements of $\mathcal{F}$. The function obeys Kolmogorov's axioms. It can be used to define probability.

3.2 Borel Algebra

3.2.1 Topology

A topology $\mathcal{T}$ on a set X is a subset of $\mathcal{P}$ (X) such that the elements of X are open. This requires both $\emptyset$ and X to be in $\mathcal{T}$, and for $\mathcal{T}$ to be closed under union and finite intersection. The elements of $\mathcal{T}$ are called open sets. Open sets allow for the definition of nearness of points relative to its neighbour without invoking the notion of distance via a metric.

3.2.2 σ-Algebra

Let X be a set and $\mathcal{P}$ (X) be its powerset. A subset $\boldsymbol{\Sigma} \subseteq \mathcal{P}$ (X) is called a sigma algebra if it satisfies the following axioms: $\boldsymbol{\Sigma}$ is in X, closed under complementation, closed under countable union. From these axioms, we have: ⃠ is in X and from DeMorgan's Law we have closed under countable intersection.

3.2.3 Borel σ-Algebra

Before, we defined sigma algebras on regular sets X. Now, we extend by defining sigma algebras on a topology $\mathcal{T}$. $\mathcal{T}$ is already a subset of $\mathcal{P}$ (X). We can form a sigma algebra from the members of $\mathcal{T}$. This is the Borel sigma algebra $\mathcal{B}$ ($\mathcal{T}$) or $\mathcal{B}$ (X). The elements of that are Borel sets, B. Through the application of sigma algebra axioms, we can see that B can contain other elements of X not present in the sets of the topology $\mathcal{T}$.

$\mathrm{X} = \mathbb{R}, \quad \tau =$ standard topology (open intervals and their unions)

$\{1\} \notin \tau$ as single points are not present is an open set in $\mathbb{R}$

Consider the sequence of open sets $U_n = \left(1 - \frac{1}{n},\ 1 + \frac{1}{n}\right) \in \tau \quad \forall n$

Through countable intersection, we can get $\{1\}$, $\bigcap_{n=1}^{\infty} U_n = \bigcap_{n=1}^{\infty} \left(1 - \frac{1}{n},\ 1 + \frac{1}{n}\right) = \{1\}$

Via only the application of sigma algebra axioms to $\mathcal{T}$, we can form $\mathcal{B}$ ($\mathcal{T}$). This is why the Borel sigma algebra is the smallest possible sigma algebra that contains $\mathcal{T}$. Other larger sigma algebras could be form from not a subset, but an entire power set $\boldsymbol{\mathcal{P}}$ (X), or also the larger Lebesgue sigma algebra $\boldsymbol{\mathcal{L}}$ (X).

Note: even if topology contains only open sets and we derive the sigma algebra from it, through derived requirement of complements, we have closed sets in the Borel sigma algebra too.

Now, we have an underlying set X, a sigma algebra $\boldsymbol{\mathcal{B}}$ ($\boldsymbol{\mathcal{T}}$) defined on that set, and we can have a measure μ, called the Borel measure to quantify the mass/size of those measurable sets in the sigma algebra. If we take $\mathbb{R}$ as our underlying set, with standard topology, we have: $(\mathbb{R}, \mathcal{B}\ (\mathcal{T}), \mu)$ which makes up a measure space.

<u>Defining measure spaces on $\mathbb{R}$ opens us up to using the tools of Real Analysis.</u>

<u>3.3 Random Variable, *X*</u>

To apply real analysis, it would be useful to transfer the measure defined on the probability measurable space (Ω, $\boldsymbol{\mathcal{F}}$), to the real measurable space $(\mathbb{R}, \mathcal{B}\ (\mathcal{T}))$. This is done through a measurable function; a random variable, X.

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

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

The above essentially means that the things that the measurable function can transfer on the real line must live in the sigma algebra.

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

$$(\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}), P_X)$$

P_X is the push-forward measure induced by X. It is the measure from the probability space transferred to work on $\mathbb{R}$. X also encodes structure preservation and transfer between measurable space; $\mathcal{F}$ is ensured equivalent to $\mathcal{B}(\mathbb{R})$.

Interesting thing when defining a random variable, the numeric choice is arbitrary as only the structure gets transferred. If we set origin to 1000 instead of 0, all other measurements will still be relative to 1000.

Random variables allow to probe the system differently. Suppose we look at a system and X is concerned with pressure, we could define Y, which would look at temperature and generate a different distribution compared to X. However, we would need to define another sigma algebra for Y, as we need the sigma algebras to match domains according to the push-forward measure axiom.

3.4 CDF

We have several functions that can be used to model P_X when applying real analysis (Order Theory, Metric, Topology and so..). The minimum one for a faithful recreation is the CDF (Cumulative Distribution Function).

Coarse $\mathbb{R}$ can be further refined by making additional structures from it to support analysis.

Order Theory

The order structure on $\mathbb{R}$ define intervals; a < x < b, (a, b).

Topology

An open interval is a subset of $\mathbb{R}$ that has no endpoint. An interval has no internal gap. It is a fixed segment of the real line. Even at the extremes, points, and their neighbours always remain in the interval.

A topology is a collection of open sets. The standard topology on $\mathbb{R}$ is a collection of all open sets in $\mathbb{R}$, generated through the arbitrary union of the basis set (the set of open intervals in $\mathbb{R}$).

Measure Theory

The Borel sigma algebra is generated from the open sets in $\mathcal{T}_{\mathbb{R}}$. Since we can generate the open sets from the basis (open intervals), The Borel sigma algebra can originate from the open intervals.

So, we can define P_X on the intervals to define the measure on whole $\boldsymbol{\mathcal{B}}$ ($\mathbb{R}$).

Furthermore, instead of using parameters a and b to define intervals (a, b], we can use half lines. $(a, b] = (-\infty, b] \setminus (-\infty, a]$. Difference of half lines recover intervals. We can also define complements, $(a, b] = (-\infty, b] \cap (-\infty, a]^c$. We have a one parameter method to describe open intervals.

We define a function F (x) to track what we want from P_X on $\mathbb{R}$.

We can thus define CDF $F(x) = P_X((-\infty, x])$.

Out of scope of this document, but:

- π-λ theorem guarantees the uniqueness of $F(x)$ in determining P_X.
- Caratheodory's Extension Theorem guarantees existence of a measure on $\boldsymbol{\mathcal{B}}(\mathbb{R})$ consistent with the CDF defined.

3.5 PDF

When we are working on $\mathbb{R}$, we now have two measures, P_X and the Lebesgue measure λ, both defined on the same measurable space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. What if we want to express a measure as a weighted version of another? This allows us to "rescale" P_X wrt the natural measure of $\mathbb{R}$.

The Radon-Nikodym Theorem state that if P_X is absolutely continuous wrt λ, we are guaranteed a unique function f such that $P_X(B) = \int_B f \; d\lambda, \quad \forall B \in \mathcal{B}(\mathbb{R})$. f is the Radon-Nikodym derivative written as $f = \frac{d\lambda}{dP_X}$. The function f is the pdf.

Relating this to the CDF:

We know $\lambda\big((a,b]\big) = b - a$; dt means integration wrt λ.

$$F(x) = P_X\big((-\infty, x]\big) = \int_{-\infty}^{x} f(t) \; d\lambda(t) = \int_{-\infty}^{x} f(t) \; dt \qquad \frac{dF}{dx} = f(x)$$

The PDF is the derivative of the CDF.

3.6 Measures of Central Tendency & Dispersion

Now we have Lebesgue measure and integral, we can obtain the expectation and variance:

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

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

4. Random Vectors

Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a probability space. A random vector in $\mathbb{R}^{\mathrm{m}}$ is a measurable map.

$$X: \Omega \to \mathbb{R}^{\mathrm{m}}$$

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

Equivalently, this means we have component random variables

$$X_i: (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$$

for $i = 1, \dots, n$, all defined on the same measurable space $(\Omega, \mathcal{F})$, and we form

$$X(\omega) = (X_1(\omega), \dots, X_n(\omega)).$$

If we want to form a random vector in the standard sense, all components must share the same domain Ω and the same σ-algebra $\mathcal{F}$. Otherwise, the tuple $(X_1, \dots, X_n)$ is not even a well-defined function on a single space.

If they are defined on different spaces**:**
Suppose

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

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

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

Then define

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

which is a valid random vector.

Same domain $(\Omega, \mathcal{F})$: directly forms a random vector.

Different domains: must embed into a common measurable space (usually a product space) before forming the vector.

Additional Stuff

12.9 Example *independent events: product probability space*

Suppose $(\Omega_1, \mathcal{F}_1, P_1)$ and $(\Omega_2, \mathcal{F}_2, P_2)$ are probability spaces. Then

$$(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2, P_1 \times P_2),$$

as defined in Chapter 5, is also a probability space.

If $A \in \mathcal{F}_1$ and $B \in \mathcal{F}_2$, then $(A \times \Omega_2) \cap (\Omega_1 \times B) = A \times B$. Thus

$$\begin{aligned}(P_1 \times P_2)\big((A \times \Omega_2) \cap (\Omega_1 \times B)\big) &= (P_1 \times P_2)(A \times B)\\ &= P_1(A) \cdot P_2(B)\\ &= (P_1 \times P_2)(A \times \Omega_2) \cdot (P_1 \times P_2)(\Omega_1 \times B),\end{aligned}$$

where the second equality follows from the definition of the product measure, and the third equality holds because of the definition of the product measure and because P_1 and P_2 are probability measures.

The equation above shows that the events $A \times \Omega_2$ and $\Omega_1 \times B$ are independent events in $\mathcal{F}_1 \otimes \mathcal{F}_2$.

Measure, Integration & Real Analysis, by Sheldon Axler