Exercise Class 3

Measures in Probability Theory

If we want to construct a measure, there is a simple and yet very fundamental procedure. We proved that any pre-measure on an algebra $\mathcal{A}$ can be extended to a measure.

1 The Extension Theorem

We start by recalling the necessary definitions.

Definition 1. Let $\mathcal{A} \subseteq \mathcal{P}(X)$ be an algebra. A mapping $\lambda : \mathcal{A} \to [0, \infty]$ is called a *pre-measure* if

i) $\lambda(\varnothing) = 0$.

ii) $\lambda(A) = \sum_{k=1}^{\infty} \lambda(A_k)$ for every $A \in \mathcal{A}$ such that $A = \bigcup_{k=1}^{\infty} A_k$, $A_k \in \mathcal{A}$ mutually disjoint.

λ is called *σ-finite* if there exists a covering $X = \bigcup_{k=1}^{\infty} S_k$, $S_k \in \mathcal{A}$ and $\lambda(S_k) < \infty$ for all k.

Remark. Infinite measures can have somewhat pathological properties at times. The σ-finiteness property is a very convenient property of (pre-)measures as it tends to share a few properties with finite measures. As we will see below, a pre-measure being σ-finite ensures that the Carathéodory–Hahn extension is unique. We will also see that some theorems require σ-finiteness as hypothesis. For example both Fubini's theorem and the Radon–Nikodym theorem are stated under the assumption of σ-finiteness of the measures involved.

Moreover, any σ-finite measure is equivalent to a probability measure.

In the special case $X = \mathbb{R}$, one can describe σ-finite pre-measures in an alternative way. This is especially useful in probability theory.

Definition 2. A *distribution function* on $\mathbb{R}$ (sometimes abbreviated as CDF, where C stands for cumulative) is a mapping $F : \mathbb{R} \to [0, \infty)$ such that

i) F is increasing, i.e., $F(x) \leq F(y)$ for $x \leq y$.

ii) F is right-continuous, i.e., $F(x_n) \to F(x)$ as $x_n \searrow x$.

iii) $\lim_{x \to -\infty} F(x) = 0$.

Proposition 3. *Let $X = \mathbb{R}$ and*

$$\mathcal{A} = \{\textit{all finite unions of } (a, b] \textit{ with } -\infty < a \leq b < \infty\}.$$

Then there exists a bijection between distribution functions on $\mathbb{R}$ and pre-measures μ on $\mathcal{A}$ with the properties

a) $0 \leq \mu\big((x, y]\big) < \infty$ *for all $x, y \in \mathbb{R}$ with $x \leq y$.*

b) $\lim_{x \to -\infty} \mu\big((x, 0]\big) < \infty$.

Proof idea. If we set $F(y) = \lim_{x \to -\infty} \mu((x, y])$ and $\mu((x, y]) = F(y) - F(x)$, then the properties of F or μ respectively follow. □

Given a pre-measure, we can now easily define a measure using the Carathéodory–Hahn extension theorem.

Theorem 4 (Carathéodory–Hahn extension)**.** *Let $\lambda : \mathcal{A} \to [0, \infty]$ be a pre-measure on X. For $A \subseteq X$, define*

$$\mu(A) = \inf\left\{ \sum_{k=1}^{\infty} \lambda(A_k) : A \subseteq \bigcup_{k=1}^{\infty} A_k, A_k \in \mathcal{A} \right\}.$$

Then μ is a measure extending λ, i.e., $\mu(A) = \lambda(A)$ for all $A \in \mathcal{A}$. Moreover, every $A \in \mathcal{A}$ is μ-measurable. If λ is σ-finite, the extension is unique (and also σ-finite).

Before we construct the Lebesgue measure, we consider a whole class of examples on $\mathbb{R}$.

Example 5. Take $X = \mathbb{R}$, $\mathcal{A} = \{\text{all finite unions of } (a, b] \text{ with } -\infty < a \leq b < \infty\}$ and a continuous function $f : \mathbb{R} \to [0, \infty)$ (the function f is often called density or pdf). Then $\mathcal{A}$ is an algebra (see Exercise 3.3) and it is easy to check that

$$\mu_f\big((a, b]\big) = \int_a^b f(x)dx$$

is a pre-measure (for the moment we can take the Riemann integral). By the previous proposition, there exists an associated distribution function F_f given by

$$F_f(b) - F_f(a) = \mu_f\big((a, b]\big) = \int_a^b f(x)dx.$$

In particular, we have $\mu_f(x) = 0$ for all $x \in \mathbb{R}$. The Carathéodory–Hahn extension then gives us a measure associated to F which extends μ_f.

The previous example together with Proposition 3 shows us the natural link between measures and probability distributions. Note however, that not every distribution is of the above form (see e.g. Cantor distribution). That is, there doesn't always exist such a density function f.

2 The Lebesgue Measure

According to Theorem 4, we only need to define a suitable algebra and pre-measure to construct the Lebesgue measure. To achieve that the Lebesgue measure is the extension of volume of the so-called elementary sets, we define the algebra of elementary sets

$$\mathcal{A} = \{A \subset \mathbb{R}^n : A \text{ is the union of finitely many disjoint intervals}\}.$$

The volume of an interval is then defined to be

$$\mathrm{vol}\big((a, b)\big) = \begin{cases} \prod_{i=1}^n (b_i - a_i) & \text{if } a_i < b_i \text{ for } 1 \leq i \leq n, \\ 0 & \text{otherwise.} \end{cases}$$

Proposition 6. *The collection $\mathcal{A}$ defines an algebra and the volume function* vol *is a pre-measure on $\mathcal{A}$.*

Proof. Exercise 3.3. □

Thanks to this proposition, the Carathéodory–Hahn extension gives us the extension of the volume function defined by

$$\begin{aligned}\mathcal{L}^n(A) &= \inf\left\{\sum_{k=1}^{\infty}\mathrm{vol}(I_k) : I_k \text{ are elementary sets and } A \subseteq \bigcup_{k=1}^{\infty}\right\}\\ &= \inf\left\{\sum_{k=1}^{\infty}\mathrm{vol}(J_k) : J_k \text{ are intervals and } A \subseteq \bigcup_{k=1}^{\infty} J_k\right\}.\end{aligned}$$

As it turns out, the Lebesgue measure $\mathcal{L}$ is the only map from Lebesgue measurable sets to $[0,\infty]$ that obeys the following axioims:

i) (Empty set) $\mathcal{L}(\varnothing) = 0$

ii) (σ-additivity): If $A_1, A_2, \dots \subset \mathbb{R}^n$ is a countable sequence of disjoint Lebesgue measurable sets, then $\mathcal{L}(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty}\mathcal{L}(A_n)$.

iii) (Translation invariance): If E is Lebesgue measurable and $x \in \mathbb{R}^d$, then $\mathcal{L}(E+x) = \mathcal{L}(E)$.

iv) (Normalization): $\mathcal{L}([0,1]^d) = 1$.

The Carathédory–Hahn extension together with Proposition 6 implies in particular the existence of the Lebesgue measure. This is in turn sufficient (in fact even equivalent) to the existence of a sequence of i.i.d. random variables with a given distribution on $\mathbb{R}$. This can then be used to show the existence of the Wiener measure on the Borel σ-algebra of $C([0,1])$, or equivalently, the existence of Brownian motion; one of the most well-known stochastic processes. Thus one can see that the existence of the Lebesgue measure has far-reaching consequences beyond the realm of measure theory itself. So how does the existence of the Lebesgue measure imply the existence of i.i.d. random variables?

The Lebesgue measure on $[0,1]$ is a probability measure (i.e., it is a normalized measure ($\lambda([0,1]) = 1$). In particular, if one chooses a point X according to this probability measure and writes it in its dyadic decomposition $X = \sum_{n\geq 1} b_n 2^{-n}$, one gets an i.i.d. sequence of fair Bernoulli random variables (b_n). That is, $P(b_n = 1) = P(b_n = 0) = 1/2$. By using the usual bijection between $\mathbb{N}$ and $\mathbb{N}^2$, we therefore get a family $(b_{i,j})_{i,j\in\mathbb{N}}$ of i.i.d. fair Bernoulli random variables and therefore the family $(Y_i)_{i\in\mathbb{N}}$ defined by $Y_i := \sum_j b_{i,j}2^{-j}$ is an i.i.d. family of uniform random variables in $[0,1]$. By choosing an appropriate function F, we get that $(F(Y_i))_{i\in\mathbb{N}}$ is a family of independent random variables according to a given distribution.