Lecture 1 8/24/2023

Outline

1) Syllabus

2) Course goals

3) Measure theory basics

Measure theory basics

Measure theory is a rigorous grounding for probability theory [subject of 205A]

Simplifies notation & clarifies concepts, especially around integration & conditioning [Pset 0]

Given a set $\mathcal{X}$, a measure μ maps subsets $A \subseteq \mathcal{X}$ to non-negative numbers $\mu(A) \in [0, \infty]$

Example $\mathcal{X}$ countable (e.g. $\mathcal{X} = \mathbb{Z}$)

Counting measure $\#(A) = \#$ points in A

Example $\mathcal{X} = \mathbb{R}^n$

Lebesgue measure

$$\lambda(A) = \int \cdots \int_A dx_1 \cdots dx_n = \mathrm{Volume}(A)$$

Standard Gaussian distribution:

$$P_Z(A) = \mathbb{P}(Z \in A) \quad \text{where } Z \sim N(0, 1)$$

$$= \int_A \phi(x)\, dx \qquad \phi(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}$$

NB Because of pathological sets, $\lambda(A)$ can only be defined for certain subsets $A \subseteq \mathbb{R}^n$ [HW 0, Prob. 3]

In general, the domain of a measure μ is a collection of subsets $\mathcal{F} \subseteq 2^{\mathcal{X}}$ (power set)

$\mathcal{F}$ must be a <u>σ-field</u> meaning it satisfies certain closure properties (not important for us)

1. $\mathcal{X} \in \mathcal{F}$
2. If $A \in \mathcal{F}$ then $\mathcal{X} \setminus A \in \mathcal{F}$
3. If $A_1, A_2, \ldots \in \mathcal{F}$ then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

<u>Ex</u>: $\mathcal{X}$ countable, $\mathcal{F} = 2^{\mathcal{X}}$

<u>Ex</u>: $\mathcal{X} = \mathbb{R}^n$, $\mathcal{F}$ = Borel σ-field $\mathcal{B}$

$\mathcal{B}$ = smallest σ-field including all open rectangles

$$(a_1, b_1) \times \cdots \times (a_n, b_n) \qquad a_i < b_i \quad \forall i$$

Given a <u>measurable space</u> $(\mathcal{X}, \mathcal{F})$ a <u>measure</u> is a map $\mu : \mathcal{F} \to [0, \infty]$ with

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i) \quad \text{for disjoint } A_1, A_2, \ldots \in \mathcal{F}$$

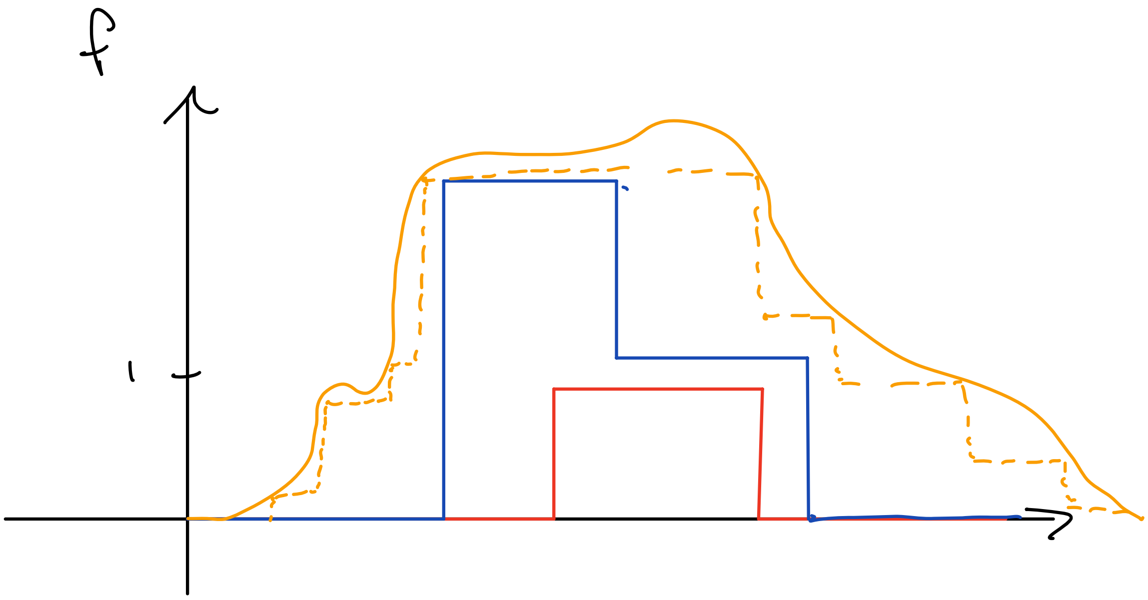
$$\mu(\emptyset) = 0$$

μ <u>probability measure</u> if $\mu(\mathcal{X}) = 1$

Integrals

Measures let us define __integrals__ that put weight $\mu(A)$ on $A \subseteq \mathcal{X}$

Define $\int 1\{x \in A\} \, d\mu(x) = \mu(A)$, extend to other functions by linearity & limits:



Indicator $\int 1\{x \in A\} \, d\mu(x) = \mu(A)$

Simple Function $\int \left(\sum c_i 1\{x \in A_i\} \right) d\mu(x) = \sum c_i \mu(A_i)$

"Nice enough" (measurable) function $\int f(x) \, d\mu(x)$ approximated by simple functions

Examples:

Counting: $\int f \, d\# = \sum_{x \in \mathcal{X}} f(x)$ ↙ Lebesgue integral

Lebesgue: $\int f \, d\lambda = \int \cdots \int f(x) \, dx_1 \cdots dx_n$

Gaussian: Note $\int 1_A(x) \, dP_Z(x) = P_Z(A) = \int_{-\infty}^{\infty} 1_A \phi \, dx$

By extension,

$$\int f \, dP_Z = \int f(x) \phi(x) \, dx = \mathbb{E}[f(Z)]$$

To evaluate $\int f \, dP_Z$ rewrite as $\int f \phi \, dx$. ↙ density [can't always do this] e.g. Binom

It is nice to turn integrals we care about into Lebesgue integrals. When can we do this?

Densities

λ and P above are closely related. Want to make this precise.

Given $(\mathcal{X}, \mathcal{F})$, two measures P, μ

We say P is <u>absolutely continuous</u> wrt μ if $P(A) = 0$ whenever $\mu(A) = 0$

Notation: <u>$P \ll \mu$</u> or we say μ <u>dominates</u> P

If $P \ll \mu$ then (under mild conditions) we can always define a <u>density function</u>

$\rho : \mathcal{X} \to [0, \infty)$ with

$$P(A) = \int_A \rho(x)\, d\mu(x)$$

$$\int f(x)\, dP(x) = \int f(x) \rho(x)\, d\mu(x)$$

Sometimes written $\rho(x) = \frac{dP}{d\mu}(x)$, called <u>Radon - Nikodym derivative</u>

Densities are very useful:

Turn $\int f(x)dP(x)$ into something we know how to evaluate, such as

1) $\int_{\mathcal{X}} f(x)p(x)dx$ (X continuous, $\mathcal{X} \subseteq \mathbb{R}^n$)

$p(x)$ called probability density function (pdf)

2) $\sum_{x \in \mathcal{X}} f(x)p(x)$ (X discrete, $\mathcal{X}$ countable)

$p(x)$ called probability mass function (pmf)

Often define distributions by giving their density wrt some known measure, e.g.

Ex: Binom(n, θ) pmf : $p(x) = \theta^n (1-\theta)^{n-x} \binom{n}{x}$, $x = 0, \ldots, n$

(density p wrt counting measure on $\mathcal{X} = \{0, \ldots, n\}$)

Note this dist. has no density wrt Lebesgue:

$\int_{\{0,\ldots,n\}} p(x)dx = 0$ for any function p

Probability Space, Random Variables

Typically, we set up a problem with multiple random variables having various relationships to one another.

Want to be able to talk about the "prob. that something happens"

Convenient setup:

R.V.s as functions of an abstract "outcome" ω

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a <u>probability space</u>

$\omega \in \Omega$ called <u>outcome</u>

$A \in \mathcal{F}$ called <u>event</u>

$\mathbb{P}(A)$ called <u>probability of A</u>

A <u>random variable</u> is a function $X : \Omega \to \mathcal{X}$

We say X has distribution Q $(X \sim Q)$ if

$$\mathbb{P}(X \in B) = \mathbb{P}(\{\omega : X(\omega) \in B\}) = Q(B)$$

More generally, could write events involving many R.V.s:

$$\mathbb{P}(X > Y > Z \geq 0) = \mathbb{P}(\{\omega : \cdots\})$$

The *expectation* is an integral w.r.t. $\mathbb{P}$

$$\mathbb{E}[f(X,Y)] = \int_{\Omega} f(X(\omega), Y(\omega))\, d\mathbb{P}(\omega)$$

To do real calculations we must eventually boil $\mathbb{P}$ or $\mathbb{E}$ down to concrete integrals/sums/etc.

If $\mathbb{P}(A) = 1$ we say A occurs *almost surely*

More in Keener ch. 1, much more in Stat 205A