Lecture 1 8/27/2020

Outline

1) Syllabus
2) 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

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 $\lambda(A) = \int \cdots \int_A dx_1 \cdots dx_n$

$= \text{Volume}(A)$

Standard Gaussian distribution:

$$P(A) = \int \cdots \int_A \phi(x)\, dx_1 \cdots dx_n \qquad \phi(x) = \frac{e^{-\frac{1}{2}\sum x_i^2}}{(2\pi)^{n/2}}$$

$= \mathbb{P}(Z \in A)$ where $Z \sim N_n(0, I_n)$

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)

<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

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

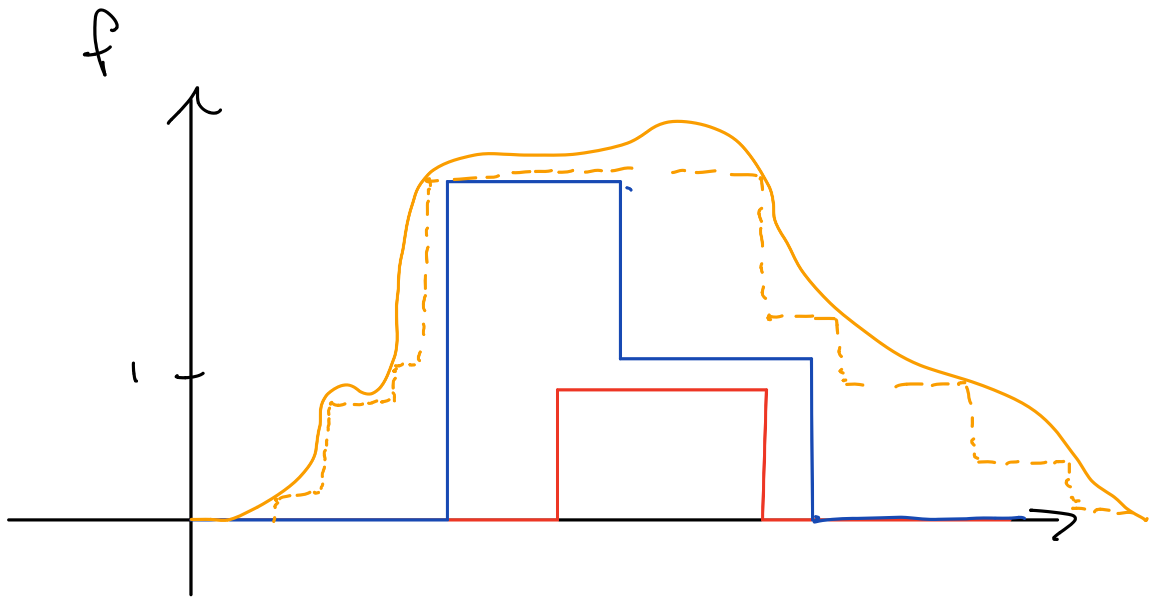
Given a <u>measurable space</u> $(\mathcal{X}, \mathcal{F})$ a <u>measure</u> is a map $\mu : \mathcal{F} \to [0, \infty]$ with

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

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

Measures let us define <u>integrals</u> 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: $\int f\, d\lambda = \int \cdots \int f(x)\, dx_1 \cdots dx_n$

Gaussian: $\int f\, dP = \int \cdots \int f(x)\, \phi(x)\, dx_1 \cdots dx_n = \mathbb{E}[f(Z)]$

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: $\underline{P \ll \mu}$ 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>

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

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

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

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

Useful to turn $\int f\, dP$ into $\int f p\, d\mu$ if we know how to calculate integrals $d\mu$

If P prob., μ Lebesgue:

$p(x)$ called <u>prob. density fcn</u> (pdf)

If P prob., μ counting:

$p(x)$ called <u>prob. mass fcn</u> (pmf)

Probability Space, Random Variables

Typically, we set up a problem with multiple random variables having various relationships to one another, convenient to think of them as functions of an abstract "outcome" ω

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

$\omega \in \Omega$ called outcome

$A \in \mathcal{F}$ called event

$\mathbb{P}(A)$ called probability of A

A random variable 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