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 \( \mu \) maps subsets \( A \subseteq \mathcal{X} \) to non-negative numbers \( \mu(A) \in [0, \infty] \)

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

\[ \text{Counting measure } \#(A) = \# \text{ points in } A \]

Example \( \mathcal{X} = \mathbb{R}^n \)

\[ \text{Lebesgue measure } \lambda(A) = \int_A \cdots dx, \cdots dx = \text{Volume}(A) \]

Standard Gaussian distribution:

\[ P_z(A) = \int A(z \in A) \quad \text{where } Z \sim \mathcal{N}(0, 1) \]

\[ = \int_A \phi(x) dx \quad \phi(x) = e^{-x^2/2} \sqrt{2\pi} \]

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 $\mu$ is a collection of subsets $\mathcal{F} \subseteq 2^X$ (power set) that must be a $\sigma$-field meaning it satisfies certain closure properties (not important for us)

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

\[ \mathcal{F} \text{ countable, } \mathcal{F} = 2^X \]
\[ \mathcal{F} = \mathbb{R}^n, \mathcal{F} = \text{Borel } \sigma\text{-field } \mathcal{B} \]
\[ \mathcal{B} \text{ is smallest } \sigma\text{-field including all open rectangles } (a_1, b_1) \times \cdots \times (a_n, b_n), a_i < b_i \forall i \]

Given a measurable space $(X, \mathcal{F})$ a measure is a map $\mu : \mathcal{F} \to [0, \infty]$ with
\[ \mu(\bigcup_{i=1}^\infty A_i) = \sum_{i=1}^\infty \mu(A_i) \text{ for disjoint } A_1, A_2, \ldots \in \mathcal{F} \]
\[ \mu(\emptyset) = 0 \]

$\mu$ is a probability measure if $\mu(X) = 1$
Measures let us define integrals that put weight $m(A)$ on $A \subseteq \mathcal{X}$

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

Indicator

$\int 1\{x \in A\} \, dm(x) = m(A)$

Simple Function

$\int \left( \sum c_i 1\{x \in A_i\} \right) \, dm(x) = \sum c_i m(A_i)$

"Nice enough" (measurable) function

$\int f(x) \, dm(x)$ approximated by simple functions
Examples:

**Counting:** \[ \sum_{x \in X} f(x) \]

**Lebesgue:** \[ \int f \, d\lambda = \int f(x) \, dx \]

**Gaussian:** Note \[ \int \mathbb{1}_A(x) \, dP_z(x) = P_z(A) = \int \mathbb{1}_A \phi \, dx \]

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. Bin]**

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

\( \lambda \) and \( P \) above are closely related. Want to make this precise.

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

we say \( P \) is absolutely continuous wrt \( \mu \)

if \( P(A) = 0 \) whenever \( \mu(A) = 0 \)

Notation: \( P \ll \mu \) or we say \( \mu \) dominates \( P \)

If \( P \ll \mu \) then (under mild conditions) we can always define a density function \( \rho : X \to [0, \infty) \) 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 Radon-Nikodym derivative
Densities are very useful:

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

1) \( \int f(x)p(x)dx \quad (\text{X continuous, } X \subseteq \mathbb{R}^n) \)

\( p(x) \) called \underline{probability density function (pdf)}

2) \( \sum_{x \in X} f(x)p(x) \quad (\text{X discrete, } X \text{ countable}) \)

\( p(x) \) called \underline{probability mass function (pmf)}

Often define distributions by giving their density \( \text{wrt some known measure, e.g.} \)

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

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

\( \text{Note this dist. has no density \( \text{wrt Lebesgue:} \) } \int p(x)dx = 0 \ \text{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" \( \omega \)

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 \( X: \Omega \to \mathbb{X} \)

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:

\[ P( X > Y > Z > 0 ) = P( \{ \omega : \cdots \} ) \]

The expectation is an integral w.r.t. \( P \)

\[ \mathbb{E}[f(X,Y)] = \int \int f(X(\omega),Y(\omega)) \, dP(\omega) \]

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

If \( P(A) = 1 \) we say \( A \) occurs \textit{almost surely}

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