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 Ω, a measure μ maps subsets Ω ⊆ Ω to non-negative numbers μ(Ω) ∈ [0, ∞]

Example Ω countable (e.g. Ω = ℤ)

Counting measure #(Ω) = # points in Ω

Example Ω = ℜ^n

Lebesgue measure λ(Ω) = ∫ₐ⁻ⁿ ∫ₐ⁻ⁿ dx₁,...,dxₙ

= Volume(Ω)

Standard Gaussian distribution:

P(Ω) = P(Z ∈ A) where Z ~ N(0, 1)

= ∫ₐ⁻ⁿ ∫ₐ⁻ⁿ φ(x) dx

φ(x) = d₋x²/₂

NB Because of pathological sets, λ(Ω) can only be defined for certain subsets Ω ⊆ 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). $\mathcal{F}$ 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 $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

Ex: $X$ countable, $\mathcal{F} = 2^X$

Ex: $X = \mathbb{R}^n$, $\mathcal{F} = \text{Borel } \sigma$-field $\mathcal{B}$

$\mathcal{B}$ is smallest $\sigma$-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

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

$\mu(\emptyset) = 0$

$\mu$ probability measure if $\mu(X) = 1$
Integrals

Measures let us define integrals that put weight $m(A)$ on $A \subseteq \mathcal{X}$.

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

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

Simple Function $\int (\sum c_i 1_{x \in A_i}) \, d\mu(x) = \sum c_i m(A_i)$

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

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

**Lebesgue:** \( \int f \, dx = \int f(x) \, dx, \ldots \, dx_n \)

**Gaussian:** Note \( \int_{-\infty}^{\infty} 1_A(x) \, d\mathcal{P}_Z(x) = \mathcal{P}(A) = \int_{-\infty}^{\infty} 1_A \phi \, dx \)

By extension,

\[
\int f \, d\mathcal{P}_Z = \int f(x) \phi(x) \, dx = \mathbb{E}[f(Z)]
\]

To evaluate \( \int f \, d\mathcal{P}_Z \) rewrite as \( \int f \phi \, dx \). [\( \phi \) density, can’t always do this.]

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

We say \(P\) is absolutely continuous wrt \(m\)

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

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

If \(P \ll m\) then (under mild conditions) we can always define a density function

\[ p : X \to [0, \infty) \quad \text{with} \]

\[ P(A) = \int_A p(x) \, dm(x) \]

\[ \int f(x) \, dP(x) = \int f(x) p(x) \, dm(x) \]

Sometimes written \( p(x) = \frac{dP}{dm}(x) \), called Radon-Nikodym derivative
Densities are very useful:

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

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

\( \rho(x) \) called probability density function (pdf)

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

\( \rho(x) \) called probability mass function (pmf)

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

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

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

Note this dist. has no density wrt Lebesgue:

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

\[ \mathbb{P}(X \in B) = \mathbb{P} \left( \{ \omega : X(\omega) \in B \} \right) = Q(B) \]
More generally, could write events involving many R.V.s:

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

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

\[ \mathbb{E}[f(x,y)] = \int f(x(\omega), y(\omega)) \, d\Pr(\omega) \]

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

If \(\Pr(A) = 1\) we say \(A\) occurs almost surely.

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