Course Review

① Tools (language/nuts & bolts of probability)

② Models for common processes

③ Applications to inference

Goals:

① Work precisely within framework of probability (following the rules)

② Be able to model real-life problems within a probabilistic framework & draw interesting conclusions (applying the rules)

Module ① Tools

Everything starts w/ Kolmogorov axioms

(Ω, F, P) = model of the universe

$P(A) \geq 0$

$P(\Omega) = 1$

$\sum_{i \geq 1} P(A_i) = P(\bigcup_{i \geq 1} A_i)$ $\quad A_1, A_2, \ldots,$ disjoint

everything in the course follows logically from this + well-known definitions

⇒ From these, we get the Law of Total Probability, Bayes rule, inclusion/exclusion, etc.

Random Variables: $X : \Omega \to \mathbb{R}$ model experiments/outcomes

↳ Described, in part, by their distribution (frequency with which different outcomes occur)

represented by: $F_X(x) = P\{X \leq x\}$, $M_X(t) = \mathbb{E}[e^{tX}]$

You can go a long way by knowing a few basic distribution:

Discrete: Bernoulli, Binomial, Geometric, Poisson

↳ coin flips, ↳ # heads, ↳ # tries, ↳ # arrivals at bus stop

Continuous: Uniform, Exponential, Gaussian

↳ time we need to wait (memoryless), ↳ fluctuation around nominal value (noise)

~ building blocks on which this course was built

Tools for working with RVs:

- Main tool: Expectation $\mathbb{E}[X] = \underbrace{\sum x\, P(x)}_{\text{discrete}} \quad \left(= \underbrace{\int x\, f(x)\,dx}_{\text{cts.}}\right)$

- other tools: clever decomposition (e.g. indicator RVs)

From this, we can compute a lot of things ($\mathbb{E} \to$ moments, variance, etc.)

Properties of $\mathbb{E}$

- linear $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$
- Tower property $\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X \mid Y]]$

ex: Suppose buses carry $Bin(n,p)$ people/bus & arrive $\sim PP(1)$. How many people arrive at the stop in 1hr, on avg?

$T = \sum_{i=1}^{N} X_i$ $\quad X_i$ = # of people on bus i

$\quad N$ = # of buses in 1 hr

$$\mathbb{E}[T] = \mathbb{E}[\mathbb{E}[T \mid N]] = \mathbb{E}[\mathbb{E}[\underbrace{\textstyle\sum_{i=1}^{N} X_i \mid N}_{np}]] = np\, \underbrace{\mathbb{E}[N]}_{1} = np$$

Oftentimes, we can't compute probabilities exactly.

Tools: concentration inequalities (Markov, Chebyshev, Chernoff)
↳ Application: information theory

Module ② $\underbrace{\text{Stochastic Processes}}_{\text{sequence of random variables}}$ & Applications

IID sequences: simplest model

$S_n = X_1 + \cdots + X_n \quad n \geq 1$

- WLLN: $\frac{1}{n} S_n \xrightarrow{\text{i.p.}} \mathbb{E}[X]$ w/ overwhelming prob
- SLLN: $\frac{1}{n} S_n \xrightarrow{\text{a.s.}} \mathbb{E}[X]$
- CLT: $\frac{1}{\sqrt{n}}(S_n - n\mathbb{E}[X_1]) \xrightarrow{\text{i.d.}} \mathcal{N}(0, var(X_1))$

3 different modes of convergence

More sophisticated models:

- DTMC : P matrix describes dynamics
  - ↝ canonical DT model for processes with "state"
    - & future depends only on current state (e.g. FSM)
    - . we can say a lot about the behavior of a DTMC
      - long-term behavior:
        - convergence to steady state/SD
        - hitting probabilities (using 1st step analysis)
        - hitting times

Poisson Processes

- canonical arrival process (uniquely defined by memoryless interarrival times)
  - Tools: merging, splitting, conditioning on arrivals
- CTMCs: described by Q-matrix
  - ↝ we can do everything we did with CTMCs
- Random Graphs
  - ↝ sharp thresholding

Module ③ Application

* - Hypothesis testing
  - likelihood ratio
  - choose desired type I error rate
* Linear estimation ⇒ optimal under Gaussian assumptions
  - ↳ *orthogonality principle *
  - ↳ application: Kalman filter