UC Berkeley
Department of Electrical Engineering and Computer Sciences

EECS 126: Probability and Random Processes

Midterm 2 Study Guide
Spring 2022

This is a brief summary of the topics we covered that will be in scope for Midterm 2. However, the scope of the exam will encompass all of the homework, discussions, lab, and lecture material; if something is not in this document, it is not necessarily out of scope. Students are expected to understand topics in more depth than they are discussed here.

1 Convergence

1. Almost sure convergence: $X_n \xrightarrow[n\to\infty]{\text{a.s.}} X$ if $\mathbb{P}(\lim_{n\to\infty} X_n = X) = 1$, i.e. the sequence $X^{(n)}$ deviates only a finite number of times from X
   (a) Strong Law of Large Numbers (empirical mean converges to true mean almost surely), convergence to stationary dist of irreducible aperiodic DTMC
   (b) Note a synonym for "almost surely" is "with probability one"
2. Convergence in probability: $X_n \xrightarrow[n\to\infty]{\text{i.p.}} X$ if $\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \epsilon) = 0$, i.e. the probability that X_n deviates only from X goes to zero (but can still deviate infinitely)
   (a) Weak Law of Large Numbers (empirical mean converges to true mean in probability)
3. Convergence in distribution: $X_n \xrightarrow[n\to\infty]{\text{d}} X$ if for all x such that $\mathbb{P}(X = x) = 0$, we have $\mathbb{P}(X_n \le x) \xrightarrow[n\to\infty]{} \mathbb{P}(X \le x)$, i.e. X_n is modeled by the distribution X
   (a) Central Limit Theorem (distribution of outcomes converges to a standard normal), Markov Chains (state distribution converges to stationary distribution)
4. We can use CLT, bounds to generate confidence intervals

2 Information Theory

1. We measure the "surprise" of a distribution with the entropy $H(X) = -\mathbb{E}[\log p(X)]$
   (a) Chain rule for entropy: $H(X,Y) = H(X) + H(Y|X)$
   (b) Mutual information: $I(X;Y) = H(X) - H(X|Y)$
2. Huffman encoding: how to construct tree, can't compress X in less than $H(X)$ bits
3. We can send information through a channel up to the capacity $C = \max_{p_X} I(X;Y)$
   (a) Know Shannon's codebook argument, an application of the probabilistic method
   (b) Understand AEP and its use in Shannon's channel coding theorem
4. Common channel examples:
   (a) Binary erasure channel: bit erased with probability p, has capacity $C = 1 - p$
   (b) Binary symmetric channel: bit swapped with probability p, has capacity $C = 1 - H(p)$
   (c) Noisy typewriter: letter shifted with probability $\frac{1}{2}$, has capacity $C = \log_2 13$

3 Discrete-Time Markov Chains

1. Markov chains satisfy the Markov property: $\mathbb{P}(X_n|X_{n-1}, X_{n-2}, ...) = \mathbb{P}(X_n|X_{n-1})$
2. Identify recurrence (positive, null), transience, irreducibility, periodicity, reversibility
3. Solving Markov chains: stationary distribution ($\pi = \pi P$), first step equations, detailed balance equations
4. Big theorem, stationary distribution, balance equations:
   (a) Detailed (a.k.a. local) balance equations hold if the Markov chain as a tree structure
   (b) Flow-in/flow-out holds for any cut, extends detailed balance equations
   (c) Stationary distribution exists for a class iff it is positive recurrent; if it exists, the stationary distribution for a communicating class is unique
   (d) If the whole chain is irreducible, then there is a unique stationary distribution
   (e) If whole chain is also aperiodic, then the chain converges a.s. to the stationary distribution for any initial distribution
5. Be able to handle an infinite number of states if necessary (ex. queue)
6. MCMC: when a probability function is intractable, we can set up a MC and sample from the stationary distribution as a proxy for sampling from the original distribution

4 Poisson Processes

1. Understand what a Poisson process is, memorylessness, independence of non-overlapping intervals; the number of arrivals in an interval of length t is distributed as Poisson(λt)
2. Distribution of arrival times, relationship between Poisson and exponential
   (a) $T_k \sim \text{Erlang}(k, \lambda)$ is the distribution of sum of k independent exponentials with rate λ
   (b) Conditioned on n arrivals up to a certain time t ($N_t = n$), $T_1, ..., T_n$ are distributed according to the order statistics of n $U[0,t]$ random variables (e.g. $\mathbb{E}[T_{i+1} - T_i] = \frac{t}{n+1}$)
3. Poisson merging: the sum of independent Poisson processes with rates λ, μ is a new Poisson process with rate $\lambda + \mu$
4. Poisson splitting: for a Poisson process with rate λ, if we label each arrival 0/1 with probability p, the 0/1 arrivals as Poisson processes with rate $p\lambda, (1-p)\lambda$ (resp.)
5. RIP: from the perspective of a given point, the expected interarrival time is doubled. A Poisson process backwards is still a Poisson process.

5 Continuous-Time Markov Chains

1. Understand how to set up rate matrix, what it means to jump states, that the "holding time" is the min of exponentials, how to calculate transition probabilities
2. Understand detailed balance equations for continuous time
3. Identify recurrence (positive, null) and transience
4. Be able to solve CTMCs for stationary distribution ($\pi Q = 0$), expected hitting times
5. Jump/embedded chain: create a DTMC that models the "jumps" of a CTMC, i.e. the visitation order of the states, by considering the transition probabilities as the min of exponentials
   (a) Transition probability from k to j is $\mathbb{P}(k, j) = \frac{\lambda_{k,j}}{\sum_{i=1}^{n} \lambda_{k,i}}$

(b) ex. modeling who will win a basketball game (first to 10 points), where teams score according to a Poisson distribution

(c) crucially, no self loops, so does not take into account holding time.

6. Uniformization: create a DTMC that has the same stationary distribution of the CTMC, by relating the rates in terms of a fixed discrete rate

    (a) Choose a fixed rate λ (we frequently use the largest sum of outgoing rates of any state in the CTMC, but any greater value also works)

    (b) Transition probability from k to j (for $k \neq j$) is $\mathbb{P}(k, j) = \frac{\lambda_{k,j}}{\lambda}$

    (c) Transition probability from k to k (self-loop) is $\mathbb{P}(k, k) = 1 - \sum_{i=1, i\neq k}^{n} \mathbb{P}(k, i)$

    (d) Also can write transition matrix P in terms of rate matrix Q as $P = I + \frac{1}{\lambda} Q$

    (e) Has the same stationary distribution: $\pi P = \pi(\frac{1}{\lambda} Q + I) = \pi(0 + I) = \pi$

6 Labs

All material from labs is in scope, but in particular you should be comfortable with the following ideas:

1. Fountain codes: familiar with the general setup and decoding scheme: problem is to design some distribution over the number of chunks in each packet

2. Matrix sketching: we want to compute $A^T B$ so we "sketch" each matrix as SA and SB for some "fat" $d \times n$ random matrix S such that $S^T S \approx I_n$.

3. The general idea of sampling from a distribution with an inhibitively large sample space by simulating a random walk on a Markov chain with the correct stationary distribution. Specifically, be familiar with the MH algorithm idea of proposing a next state and accepting with some probability.

4. Perlin Noise can be used to generate correlated noise with controlled frequency

5. Guess My Word: the idea of binary search and using Huffman encodings to "ask" as few questions as possible – equivalent to minimum compression length