UC Berkeley
Department of Electrical Engineering and Computer Sciences

EECS 126: Probability and Random Processes

Midterm 2 Study Guide
Fall 2019

This is a checklist of topics that are in scope for the midterm; students are expected to understand topics in more depth than they are discussed here.

1 Bounds/Concentration Inequalities, MGFs

1. Markov's inequality: $\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$ for nonnegative random variable X
2. Chebyshev's inequality: $\mathbb{P}(|X - \mathbb{E}[X]| \geq c) \leq \frac{\text{Var}(X)}{c^2}$
3. Moment generating functions are given by $M_x(s) = \mathbb{E}[e^{sX}]$; be able to recognize these for common distributions, and read off the parameters
4. We can combine MGFs with Markov's inequality to get Chernoff's inequality: $\mathbb{P}(X \geq a) = \mathbb{P}(e^{sX} \geq e^{sa}) \leq \frac{\mathbb{E}[e^{sX}]}{[e^{sa}]}$, and then taking the min over all s to get the best bound

2 Convergence, Law of Large Numbers, CLT

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), AEP (a random sequence falls inside the typical set almost surely)
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 $\mathbb{P}(X_n \leq x) \xrightarrow[n\to\infty]{} \mathbb{P}(X \leq x)$, 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 or bounds to set up confidence intervals for estimation

3 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) Binary erasure/symmetric channels, know Shannon's codebook argument

4 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: first step equations ($\pi = \pi P$), 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 aperiodic, then the chain converges to the stationary distribution for any initial distribution
   (f) The reciprocal of the stationary distribution is the expected time to return to a state given that you start from that state
5. Be able to handle an infinite number of states if necessary
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; understand how Metropolis Hastings can be used to solve computation problems, even if they aren't strictly probabilistic (e.g. finding the correct cipher in lab)

5 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) $S_k \sim$ Erlang(k, λ) is the distribution of sum of k independent exponentials with rate λ
   (b) Conditioned on n arrivals up to a certain time s ($N(s) = n$), $S_1, ..., S_n$ are distributed according to the order statistics of n $U[0, s]$ random variables (e.g. $\mathbb{E}[S_{i+1} - S_i] = \frac{1}{n+1}$)
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

6 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, expected hitting times
5. We can set up a "jump chain" using the transition probabilities of a CTMC, converting it into a DTMC where we care about the arrivals (but lose information about times)