UC Berkeley
Department of Electrical Engineering and Computer Sciences

EECS 126: Probability and Random Processes

Final Study Guide
Fall 2019

This is a checklist of topics that were covered after midterm 2; while these topics will be emphasized in the final, the scope of the final is the entire class. Students are expected to understand topics in more depth than they are discussed here.

1 Markov Chains and Erdos-Renyi Random Graphs

1. Remember details about DTMCs, CTMCs from before; they are still important
2. ER Graph $\mathcal{G}(n, p)$: random graph of n vertices where each edge is independently picked to be in the graph with probability p, no self-loops
3. Know how to do calculations for degree distribution, expectations, and other simple random quantities associated with the graph
4. Understand how tail bounds can help establish thresholds for random graphs

2 Estimation, MLE, MAP

1. The expected error of an estimator $\hat{y} = f(X)$ is $\mathbb{E}[f(X) - y]$
2. The bias of an estimator is $\mathbb{E}[f(X) - y]$ – when this is zero, we say the estimator is unbiased; note that minimizing bias is not the same as minimizing expected error
3. Maximum Likelihood Estimation (MLE): find parameters θ of model that maximize the likelihood $l(X|\theta)$
   (a) Most of the time, $l(X|\theta)$ has form $\prod_{i=1}^{n} f(x_i|\theta)$, so it is easier to equivalently maximize the log-likelihood, which will have form $\sum_{i=1}^{n} \log f(x_i|\theta)$
   (b) The MLE estimator can be biased, ex. German tank problem, finding variance of a Gaussian from samples
   (c) MLE is a special case of MAP, where the prior over θ is uniform
   (d) Most statistics/machine learning is MLE with a uniform prior on the dataset
4. Maximum a Posteriori Estimation (MAP): find parameters θ of model that maximize $l(X|\theta)f(\theta)$, where $f(\theta)$ is a prior probability distribution over θ
   (a) Adding a regularizer to a cost function typically corresponds to some form of MAP estimation, ex. $L(x) = \sum_{i=1}^{n}(y_i - wx_i)^2 + \lambda|w|$ corresponds to a Laplace prior
   (b) When performing statistics/machine learning on a dataset, MAP can correspond to some datapoints being more important than others

3 Hypothesis Testing

1. Neyman-Pearson hypothesis testing: a form of frequentist hypothesis testing, where we assume no prior over possible values for the parameter X we want to estimate
   (a) We suppose there are two outcomes, either $X = 0$ – the null hypothesis – or $X = 1$, the alternate hypothesis
   (b) Since we have no prior, there is no notion of the "most likely" outcome: we have to instead measure PFA, PCD, which we can measure since they assume that a given outcome is true
   (c) Probability of False Alarm (PFA): $P(\hat{X} = 1|X = 0)$
   (d) Probability of Correct Detection (PCD): $P(\hat{X} = 1|X = 1)$
   (e) Goal of N-P hypothesis testing: maximize PCD such that the PFA is less than β
   (f) ROC curve: maximizing PCD is equivalent to maximizing PFA
   (g) In order to maximize PFA, we may need to add some probability γ of reporting that $\hat{X} = 0$ on the decision boundary
   (h) Intuitively, we seek to answer in the form of a p-value: given that $X = 0/X = 1$ is true, what is the probability we observed this data?
2. Bayesian hypothesis testing: minimize an estimated cost $\mathbb{E}[L(Y)]$ for data Y, where there is some prior over the possible parameters $X = 0$ or $X = 1$

4 Vector Space of Random Variables

1. We consider a Hilbert space of random variables, where $\langle X, Y\rangle = \mathbb{E}[XY]$, with corresponding norm $\|X\|^2 = \mathbb{E}[X^2]$
2. Linear Least Squares Estimation (LLSE): we have three vectors $1, X, Y$ – we want to find Y given the data X, it terms of a linear combination of X and 1
   (a) This corresponds to finding the projection of Y onto both $\tilde{X}$ and 1, where $\tilde{X}$ is the transformed X such that $\langle \tilde{X}, 1\rangle = 0$
   (b) $L[Y|X] = \mathbb{E}[Y] + \frac{\text{cov}(X,Y)}{\text{var}(X)}(X - \mathbb{E}[X])$
   (c) If the noise is Gaussian, the LLSE is also the MMSE
3. Minimum Mean Square Estimation (MMSE): find the best function ϕ to minimize the expected squared error $\mathbb{E}(Y - \phi(X))^2]$
   (a) Here, we want $\langle Y - \phi(X), f(X)\rangle = 0$ for all other f
   (b) In particular, the MMSE estimator is $\mathbb{E}[Y|X]$
4. Both the LLSE and MMSE are unbiased, as $\mathbb{E}[X - \mathbb{E}[X]] = \mathbb{E}[\mathbb{E}[Y|X] - \mathbb{E}[Y]] = 0$

5 Kalman Filtering

1. Jointly Gaussian: two random variables X, Y are said to be jointly Gaussian if (X, Y) is Gaussian; the linear combination of Gaussians is Gaussian
2. Setup: we have some hidden, time varying parameters X_i that output noisy observations Y_i, where both sources of noise are Gaussian
3. Kalman filtering: given observations $Y_1, ..., Y_n$, a recursive set of update rules to estimate the underlying parameter X_n
4. Kalman smoothing: given observations $Y_1, ..., Y_n$, estimate the past underlying parameter X_i, for some $i < n$