UC Berkeley
Department of Electrical Engineering and Computer Sciences

EECS 126: Probability and Random Processes

Homework 11
Fall 2023

1. **Connected Random Graph**

   We start with the empty graph on n vertices. Iteratively, we add an undirected edge, chosen uniformly at random from the edges that are not yet present in the graph, until the graph is connected.

   Hint: Recall the coupon collector's problem.

   a. Suppose that there are currently k connected components in the graph. Let X_k be the number of edges we need to add until there are $k-1$ connected components. Show that $\mathbb{E}(X_k) \leq \frac{n-1}{k-1}$.
   b. Let X be the total number of edges in the final connected graph. Show that $\mathbb{E}(X) \leq Cn \log n$ for some constant C.

2. **Isolated Vertices**

   Consider an Erdős–Rényi random graph $\mathcal{G}(n, p(n))$, where n is the number of vertices and $p(n)$ is the probability that any specific edge appears in the graph. Let X_n be the number of isolated vertices in $\mathcal{G}(n, p(n))$.

   a. Show that $\mathbb{E}(X_n) \to \exp(-c)$ as $n \to \infty$ when $p(n) = \frac{(\ln n)+c}{n}$ for some constant c.
   b. Conclude that $\mathbb{E}(X_n) \to \infty$ when $p(n) \ll \frac{\ln n}{n}$.
   c. Conclude that $\mathbb{E}(X_n) \to 0$, and $X_n \to 0$ in probability, when $p(n) \gg \frac{\ln n}{n}$.

   The asymptotic notation $f(n) \ll g(n)$ means that $\frac{f(n)}{g(n)} \to 0$ as $n \to \infty$.

   Hint: From Taylor series expansion, $\ln(1 + x) \approx x$ when x is small.

3. **Community Detection Using MAP**

It may be helpful to work on this problem in conjunction with the relevant lab. The *stochastic block model* (SBM) defines the random graph $\mathcal{G}(n, p, q)$ consisting of two communities of size $\frac{n}{2}$ each, such that the probability an edge exists between two nodes of the same community is p, and the probability an edge exists between two nodes in different communities is $q < p$. The goal of the problem is to exactly determine the two communities, given only the graph.

Show that the MAP estimate of the two communities is equivalent to finding the *min-bisection* or *balanced min-cut* of the graph, the split of G into two groups of size $\frac{n}{2}$ that has the minimum edge weight across the partition. Assume that any assignment of the communities is a priori equally likely.

4. **Bayesian Estimation of Exponential Distribution**

We have seen the MLE (non-Bayesian perspective) and MAP estimation (Bayesian perspective). In this problem, we will introduce the fully Bayesian approach to statistical estimation.

Suppose that X is Exponential with unknown rate Λ. As a Bayesian practitioner, you have a prior belief that the random variable Λ is equally likely to be λ_1 or λ_2.

Now, you collect one sample X_1 from X.

a. Find the posterior distribution $\mathbb{P}(\Lambda = \lambda_1 \mid X_1 = x_1)$.

b. If we were using the MLE or MAP rule, we would choose a single value λ for Λ, sometimes called a *point estimate*. This amounts to saying X has Exponential distribution with rate λ. In the Bayesian approach, we will instead keep the full information of the posterior distribution of Λ, and we compute the distribution of X as

$$f_X(x) = \sum_{\lambda \in \{\lambda_1, \lambda_2\}} f_{X|\Lambda}(x \mid \lambda) \cdot \mathbb{P}(\Lambda = \lambda \mid X_1 = x_1).$$

Note that we do not necessarily have an Exponential distribution for X anymore. Compute $f_X(x)$ in closed form.

c. From the previous part, you may have guessed that the fully Bayesian approach is often computationally intractable, which is one of the main reasons why point estimates are common in practice. Supposing that $\lambda_1 > \lambda_2$, compute the MAP estimate for Λ, and calculate $f_X(x)$ again using the MAP rule.

5. **Linear Regression, MLE, and MAP**

Suppose you draw n i.i.d. data points $(x_1, y_1), \ldots, (x_n, y_n)$, where the true relationship is given by $Y = WX + \varepsilon$ for $\varepsilon \sim \mathcal{N}(0, \sigma^2)$. In other words, Y has a linear dependence on X with additive Gaussian noise.

a. Show that finding the MLE of W given the data points $\{(x_i, y_i)\}_{i=1}^n$ is equivalent to minimizing mean squared error, or minimizing the cost function

$$J(w) = \sum_{i=1}^{n} (y_i - wx_i)^2.$$

b. Now suppose that W has a *Laplace* prior distribution,

$$f_W(w) = \frac{1}{2\beta} e^{-|w|/\beta}.$$

Show that finding the MAP estimate of W given the data points $\{(x_i, y_i)\}_{i=1}^n$ is equivalent to minimizing the cost function

$$J(w) = \sum_{i=1}^{n} (y_i - wx_i)^2 + \lambda|w|.$$

(You should determine what λ is.) This is interpreted as a one-dimensional ℓ^1-regularized least-squares criterion, also known as LASSO.

6. **Minimum-Error Property of MAP**

a. Let $X \in \{0, 1\}$, and suppose we have the prior $\mathbb{P}(X = 0) = \pi_0$ and $\mathbb{P}(X = 1) = \pi_1$. Let $\hat{X}_{\text{MAP}}$ be the MAP estimate of X given the random variable Y, and let $\hat{X}$ be any other estimate of X given Y. Show that

$$\mathbb{P}(X \neq \hat{X}_{\text{MAP}}) \leq \mathbb{P}(X \neq \hat{X}).$$

b. Now, also suppose that type I errors (declaring $\hat{X} = 1$ when $X = 0$) incur a cost of $c_1 \geq 0$ and type II errors (declaring $\hat{X} = 0$ when $X = 1$) a cost of $c_2 \geq 0$. Derive the decision rule $\hat{X}$ that minimizes the total cost

$$c_1 \, \mathbb{P}(\hat{X} = 1, X = 0) + c_2 \, \mathbb{P}(\hat{X} = 0, X = 1).$$