UC Berkeley
Department of Electrical Engineering and Computer Sciences

EECS 126: Probability and Random Processes

Problem Set 10
Spring 2018

Issued: April 2, 2018 **Due:** 11:59 PM, Wednesday, April 11, 2018

---

1. **Midterm**

   Solve all of the problems on the midterm again (including the ones you got correct).

2. **Exponential: MLE & MAP**

   The random variable X is exponentially distributed with mean 1. Given X, the random variable Y is exponentially distributed with rate X.

   (a) Find $\text{MLE}[X \mid Y]$.

   (b) Find $\text{MAP}[X \mid Y]$.

3. **BSC: MLE & MAP**

   You are testing a digital link that corresponds to a BSC with some error probability $\epsilon \in [0, 0.5]$.

   (a) Assume you observe the input and the output of the link. How do you find the MLE of ϵ?

   (b) You are told that the inputs are i.i.d. bits that are equal to 1 with probability 0.6 and to 0 with probability 0.4. You observe n outputs (n is a positive integer). How do you calculate the MLE of ϵ?

   (c) The situation is as in the previous case, but you are told that ϵ has PDF $4 - 8x$ on $[0, 0.5)$. How do you calculate the MAP of ϵ given n outputs?

4. **Fun with Linear Regression**

   Suppose $f : \mathbb{R}^d \to \mathbb{R}$ is an unknown linear function, i.e. it is of the form $f(x) = x^\top w = x_1 w_1 + \cdots + x_d w_d$, where $w \in \mathbb{R}^d$ is the unknown parameter of the linear function. We pick n points $x^{(1)}, \ldots, x^{(n)} \in \mathbb{R}^d$, and we observe $y^{(1)}, \ldots, y^{(n)} \in \mathbb{R}$ that are generated according to the model

   $$y^{(i)} = f\big(x^{(i)}\big) + \epsilon_i, \text{ for } i = 1, \ldots, n,$$

   where $\epsilon_1, \ldots, \epsilon_n$ are i.i.d. $\mathcal{N}(0, \sigma^2)$ random variables.

   Let us first estimate w when we have **no prior** information about it.

   (a) Compute the likelihood of the parameter w given the data $\{(x^{(i)}, y^{(i)})\}_{i=1}^n$

   $$\mathcal{L}(w \mid \{(x^{(i)}, y^{(i)})\}_{i=1}^n) := \prod_{i=1}^n p\big(y^{(i)} \mid x^{(i)}; w\big).$$

(b) Explicitly define a matrix $X \in \mathbb{R}^{n \times d}$ and a vector $y \in \mathbb{R}^n$ such that the optimal points of the problem

$$\min_{w \in \mathbb{R}^d} \|Xw - y\|_2^2,$$

correspond to the maximizers of the likelihood.

Now assume a zero-mean **Gaussian prior** for each $w_i,\ i = 1, \ldots, d$. In particular assume that $w_1, \ldots, w_d$ are i.i.d. $\mathcal{N}(0, \tau^2)$, and they are also independent of the data.

(c) Compute, up to a normalization constant, the posterior distribution of w given the data $\{(x^{(i)}, y^{(i)})\}_{i=1}^n$.

(d) Explicitly define a matrix $X \in \mathbb{R}^{n \times d}$, a vector $y \in \mathbb{R}^n$ and a positive scalar $\lambda \in \mathbb{R}$ such that the optimal point of the problem

$$\min_{w \in \mathbb{R}^d} \|Xw - y\|_2^2 + \lambda \|w\|_2^2,$$

correspond to the maximizer of the posterior distribution of w.

5. Community Detection Using MAP

It will be useful to work on this problem in conjunction with Q3 of Lab 6. The stochastic block model (SBM), as defined in Lab 6 is a random graph $\mathcal{G}(n, p, q)$ consisting of two communities of size $n/2$ 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, where $p > q$. 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 of the graph (i.e. the split of G into two groups of size $n/2$ that has the minimum edge weight across the partition).

6. [Bonus] Bayesian Estimation of Poisson Distribution

The bonus question is just for fun. You are not required to submit the bonus question, but do give it a try and write down your progress.

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

Suppose that X_i, $i = 1, \ldots, n$, are i.i.d. drawn from a Poisson distribution of unknown mean M (M is a random variable). As a Bayesian practitioner, you have a prior belief that M is Erlang of order k and rate α.

(a) Find the posterior distribution $f_{M|X}(\mu \mid x_1, \ldots, x_n)$.

(b) If we were using the MLE or MAP rule, then we would choose a single value μ for M; this is sometimes called a *point estimate*. This amounts to saying X has the Poisson distribution with mean μ.

In the Bayesian approach, we will not use a point estimate. Instead, we will keep the full information of the posterior distribution of μ, and we compute the distribution of X as

$$\mathbb{P}(X = x) = \int_0^\infty \mathbb{P}(X = x \mid M = \mu) f_{M|X}(\mu \mid x_1, \ldots, x_n)\, \mathrm{d}\mu.$$

Notice that in the Bayesian approach, we do not necessarily have a Poisson distribution for X anymore. Compute $\mathbb{P}(X = x)$ in closed-form.

(c) You may have noticed from the previous part that the fully Bayesian approach is often computationally intractable. This is one of the main reasons why point estimates are common in practice.

Compute the MAP estimate for M and calculate $\mathbb{P}(X = x)$ again using the MAP rule.