UC Berkeley
Department of Electrical Engineering and Computer Sciences

EE126: Probability and Random Processes

Problem Set 11
Fall 2018

Issued: Wednesday, October 31, 2018 **Due:** Wednesday, November 7, 2018

Problem 1. **Poisson Process MAP**
Customers arrive to a store according to a Poisson process of rate 1. The store manager learns of a rumor that one of the employees is sending $1/2$ of the customers to the rival store. Refer to hypothesis $X = 1$ as the rumor being true, that one of the employees is sending every other customer arrival to the rival store and hypothesis $X = 0$ as the rumor being false, where each hypothesis is equally likely. Assume that at time 0, there is a successful sale. After that, the manager observes $S_1, S_2, \ldots, S_n$ where n is a positive integer and S_i is the time of the ith subsequent sale for $i = 1, \ldots, n$. Derive the MAP rule to determine whether the rumor was true or not.

Problem 2. **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?

Problem 3. **Bayesian Estimation of Exponential Distribution**
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 is an exponential random variable with unknown rate Λ (Λ is a random variable). As a Bayesian practitioner, you have a prior belief that Λ is equally likely to be λ_1 or λ_2.
You collect one sample X_1 from X.

1. Find the posterior distribution $\Pr(\Lambda = \lambda_1 \mid X_1 = x_1)$.

2. If we were using the MLE or MAP rule, then we would choose a single value λ for Λ; this is sometimes called a *point estimate*. This amounts to saying X has the exponential distribution with rate λ.

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

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

   Notice that in the Bayesian approach, we do not necessarily have an exponential distribution for X anymore. Compute $f_X(x)$ in closed-form.

3. You might guess 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 Λ and calculate $f_X(x)$ again using the MAP rule.

Problem 4. **Voltage MAP**
You are trying to detect whether voltage V_1 or voltage V_2 was sent over a channel with independent Gaussian noise $Z \sim N(V_3, \sigma^2)$. Assume that both voltages are equally likely to be sent.

(a) Derive the MAP detector for this channel.

(b) Using the Gaussian Q-function, determine the average error probability for the MAP detector.

(c) Suppose that the average transmit energy is $(V_1^2 + V_2^2)/2$ and that the average transmit energy is constrained such that it cannot be more than $E > 0$. What voltage levels V_1, V_2 should you choose to meet this energy constraint but still minimize the average error probability?

Problem 5. **Bonus: 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.

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

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

2. 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.

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

4. 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.