UC Berkeley
Department of Electrical Engineering and Computer Sciences

EECS 126: Probability and Random Processes

**Problem Set 11 (optional)**
Fall 2020

1. **Laplace Prior & $\ell^1$-Regularization**

   Suppose you draw $n$ i.i.d. data points $(x_1, y_1), \dots, (x_n, y_n)$, where $n$ is a positive integer and the true relationship is $Y = WX + \varepsilon$, $\varepsilon \sim \mathcal{N}(0, \sigma^2)$. (That is, $Y$ has a linear dependence on $X$, with additive Gaussian noise.) Further suppose that $W$ has a prior distribution with density

   $$f_W(w) = \frac{1}{2\beta} \mathrm{e}^{-|w|/\beta}, \qquad \beta > 0.$$

   (This is known as the **Laplace distribution**.) Show that finding the MAP estimate of $W$ given the data points $\{(x_i, y_i) : i = 1, \dots, n\}$ is equivalent to minimizing the cost function

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

   (you should determine what $\lambda$ is). This is interpreted as a one-dimensional $\ell^1$-regularized least-squares criterion, also known as LASSO.

2. **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 $S_1, S_2, \dots, S_n$ where $n$ is a positive integer and $S_i$ is the time of the $i$th subsequent sale for $i = 1, \dots, n$. Derive the MAP rule to determine whether the rumor was true or not. 5cm

3. **BSC Hypothesis Testing**

   Consider a BSC with some error probability $\epsilon \in [0.1, 0.5)$. Given $n$ inputs and outputs $(x_i, y_i)$ of the BSC, solve a hypothesis problem to detect that $\epsilon > 0.1$ with a probability of false alarm at most equal to 0.05. Assume that $n$ is very large and use the CLT.

   *Hint*: The null hypothesis is $\epsilon = 0.1$. The alternate hypothesis is $\epsilon > 0.1$, which is a **composite hypothesis** (this means that under the alternate hypothesis, the probability distribution of the observation is not completely determined; compare this to a **simple hypothesis** such as $\epsilon = 0.3$, which *does* completely determine the probability distribution of the observation). The Neyman-Pearson Lemma we learned in class applies for the case of a simple null hypothesis and a simple alternate hypothesis, so it does not directly apply here.

   To fix this, fix some specific $\epsilon' > 0.1$ and use the Neyman-Pearson Lemma to find the optimal hypothesis test for the hypotheses $\epsilon = 0.1$ vs. $\epsilon = \epsilon'$. Then, argue that the optimal decision rule does not depend on the specific choice of $\epsilon'$; thus, the decision rule you derive will be *simultaneously* optimal for testing $\epsilon = 0.1$ vs. $\epsilon = \epsilon'$ for all $\epsilon' > 0.1$.

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?

5. **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 $\Lambda$ ($\Lambda$ is a random variable). As a Bayesian practitioner, you have a prior belief that $\Lambda$ is equally likely to be $\lambda_1$ or $\lambda_2$.

   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, then we would choose a single value $\lambda$ for $\Lambda$; this is sometimes called a *point estimate*. This amounts to saying $X$ has the exponential distribution with rate $\lambda$.

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

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

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

   (c) 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 $\Lambda$ and calculate $f_X(x)$ again using the MAP rule.

6. **[Bonus] $p$-Value**

   *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.*

   Let us define what the $p$-value of a hypothesis test is. Given an observation $Y$ and a constraint of $\beta$ on the PFA, the Neyman-Pearson rule will either declare that the alternate hypothesis is true or not. The constraint on the PFA controls the trade-off between declaring the alternate hypothesis to be true when it is not (false alarm), and declaring the alternate hypothesis to be true when it is (correct detection). Therefore, for very high values of $\beta$, the hypothesis test will declare that the alternate hypothesis is true, and for very low values of $\beta$, the hypothesis test will declare that the null hypothesis is true.

(Intuitively, the smaller the value of $\beta$, the more conservative the resulting hypothesis test is, i.e., it will be more reluctant to declare that the alternate hypothesis is true.)

The $p$**-value** of the observation is the smallest value of $\beta$ such that the alternate hypothesis is declared true.

Think about this carefully, and explain why the $p$-value is *not* the probability that the alternate hypothesis is true.