UC Berkeley
Department of Electrical Engineering and Computer Sciences

EECS 126: PROBABILITY AND RANDOM PROCESSES

Problem Set 12
Fall 2019

1. **Laplace Prior & ℓ^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

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

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

2. **Flipping Coins and Hypothesizing**

   You flip a coin until you see heads. Let

   $$X = \begin{cases} 1 & \text{if the bias of the coin is } q > p. \\ 0 & \text{if the bias of the coin is } p. \end{cases}$$

   Find a decision rule $\hat{X}(Y)$ that maximizes $\mathbb{P}[\hat{X} = 1 \mid X = 1]$ subject to $\mathbb{P}[\hat{X} = 1 \mid X = 0] \le \beta$ for $\beta \in [0, 1]$. Remember to calculate the randomization constant γ.

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 ϵ'; thus, the decision rule you derive will be *simultaneously* optimal for testing $\epsilon = 0.1$ vs. $\epsilon = \epsilon'$ for all $\epsilon' > 0.1$.

4. **Exam Difficulties**

   The difficulty of an EECS 126 exam, Θ, is uniformly distributed on $[0, 100]$, and Alice gets a score X that is uniformly distributed on $[0, \Theta]$. Alice gets her score back and wants to estimate the difficulty of the exam.

   (a) What is the LLSE for Θ?

   (b) What is the MAP of Θ?

5. **Photodetector LLSE**

   Consider a photodetector in an optical communications system that counts the number of photons arriving during a certain interval. A user conveys information by switching a photon transmitter on or off. Assume that the probability of the transmitter being on is p. If the transmitter is on, the number of photons transmitted over the interval of interest is a Poisson random variable Θ with mean λ, and if it is off, the number of photons transmitted is 0. Unfortunately, regardless of whether or not the transmitter is on or off, photons may be detected due to "shot noise". The number N of detected shot noise photons is a Poisson random variable N with mean μ, independent of the transmitted photons. Let T be the number of transmitted photons and D be the number of detected photons. Find $L[T \mid D]$.

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 β 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 β, the hypothesis test will declare that the alternate hypothesis is true, and for very low values of β, the hypothesis test will declare that the null hypothesis is true.

(Intuitively, the smaller the value of β, 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 β 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.