MASSACHUSETTS INSTITUTE OF TECHNOLOGY

6.s077 Spring 2018

Problem Set 3. Due Tuesday 3/6, in class

Problem 1. Let Θ be equally likely to be 1 or 2. Conditioned on $\Theta = \theta$, let X be exponential with parameter θ. Find $\mathbb{P}(\Theta = 1 \mid X = x)$ (as a function of x), and find the form of the MAP estimator of Θ based on X.

Problem 2. Bayesian estimation with a prior that favors a zero value.
We make an observation $Y = \Theta + W$, where Θ and W are independent, and W is a standard normal random variable.

(a) Find the MAP estimator of Θ based on Y, for the case where our prior distribution for Θ is also a standard normal.

(b) Find the Find the MAP estimator of Θ based on Y, for the case where our prior distribution for Θ has the Laplace distribution (a.k.a. two-sided exponential) below:

$$f_\Theta(\theta) = \frac{1}{2}e^{-|\theta|}, \qquad -\infty < \theta < \infty.$$

Hint: Because $f_\Theta(\theta)$ is not differentiable at zero, you may need some more thinking, beyoind just trying to set a derivative to zero.

Discussion: If you have done part (b) correctly, you will see that the estimate will often be exactly zero. For this reason, the Laplace prior is often used in settings where we think that a parameter may be exactly equal to zero. It plays an important role in "sparse linear regression" a topic that we will visit later in this class.

Now, if one really believes that Θ might be exactly zero, one could try a prior that, for example, mandates that (i) $\mathbb{P}(\Theta = 0) = 1/2$ and (ii) conditional on $\Theta \neq 0$, Θ is a standard normal. If you are curious to work through this setting, you will find that the MAP estimator turns out to be useless: it always returns a value of zero. For this reason, if one is committed to MAP estimators, the Laplace prior better captures the intended purpose.

Optional continuation. If interested in practicing your ability to calculate, you may want to continue part (a) and obtain formulas for

$$\mathbb{E}\Big[(\hat{\Theta}_{\mathrm{MAP}} - \Theta)^2 \mid X = x\Big], \qquad \mathbb{E}\Big[(\hat{\Theta}_{\mathrm{MAP}} - \Theta)^2 \mid \theta = \theta\Big], \qquad \mathbb{E}\Big[(\hat{\Theta}_{\mathrm{MAP}} - \Theta)^2\Big].$$

Problem 3. Consider a model of vehicle motion that is more realistic than the one used in Lecture 6. Instead of letting the velocities V_t be independent, we model the fact that velocity cannot change suddenly, and consider a model of the form

$$V_t = \frac{2}{3}V_{t-1} + A_t,$$

where we model the random accelerations A_t as independent standard normal random variables. For simplicity, we assume that $V_0 = X_0 = 0$. As in Lecture, we let $X_t = X_{t-1} + V_t$, and let the measurements satisfy $Y_t = X_t + W_t$, for $t = 1, \ldots, T$. Here, each W_t is also standard normal. We assume that the random variables $W_1, \ldots, W_T, A_1, \ldots, A_T$ are all independent. For simplicity, assume that $T = 3$.

Let $\Theta = (A_1, A_2, A_3)$ and $Y = (Y_1, Y_2, Y_3)$. Express Y as a linear function of Θ and W_1, W_2, W_3, and use this to write down an expression for $f_{\Theta|Y}(\theta \,|\, y)$. When you write down your answer, you can ignore any multiplicative terms that do not involve θ.

Problem 4. A random variable Z has a distribution that is symmetric around zero, and satisfies $\mathbb{E}[Z^2] = 1$, $\mathbb{E}[Z^4] = 3$, $\mathbb{E}[Z^6] = 15$. We obtain two measurements,

$$Y_2 = Z^2 + W_2, \qquad Y_3 = Z^3 + W_3,$$

where W_2, W_3 are independent from each other and from Z, and have zero mean and unit variance.

We consider estimators of the form

$$\hat{Z} = a_2 Y_2 + a_3 Y_3.$$

Find values for a_2, a_3 such that the mean squared error $\mathbb{E}[(Z - \hat{Z})^2]$ is minimized. *Hint:* First figure out what a_2 should be.

Problem 5. This "theoretical problem" is intended to have you go though some standard but important textbook material. You are allowed to use any resources you wish in any way that you want (other than submitting a photocopy of a textbook!).

Let Y and Z be two random variables with known joint distribution. We are interested in properties of $\hat{Y} = \mathbb{E}[Y \,|\, Z]$ and of the associated error $\tilde{Y} = \hat{Y} - Y$. For simplicity, assume that Y and Z have zero mean.

(a) Show that $\mathbb{E}[\tilde{Y} \,|\, Z = z] = 0$, for any z.

(b) Show that $\mathbb{E}[\tilde{Y}] = 0$.
Comment: This implies that $\mathbb{E}[\hat{Y}] = \mathbb{E}[Y]$, and the estimator is said to be unbiased, in the Bayesian sense. Note that this is a different definition from the one used in the classical setting.

(c) Show that $\mathbb{E}[\tilde{Y}\hat{Y} \mid Z = z] = 0$, for any z.

(d) Show that $\mathbb{E}[\tilde{Y}\hat{Y}] = 0$.

(e) Show that $\mathbb{E}[Y^2] = \mathbb{E}[\hat{Y}^2] + \mathbb{E}[\tilde{Y}^2]$.

Hint: $\mathbb{E}[h(V)W \mid V] = h(V)\mathbb{E}[W \mid V]$.