ES 201 Homework 5, Due 04/13/2017 (in class)

<u>Reading</u>: Secs. 17.3, 4.10, 17.4, Lec. 19b, 20.

Problem 1

In this problem set, we will compare dynamic estimation methods to obtain filtered (sequential) estimates of the probability of spiking of a thalamic neuron in a 1 ms bins, in response to whisker deflections. The data consists of the spiking activity of a thalamic neuron in response to whisker stimulus. The whisker was displaced up and down at 8 Hz at 16 mm/second for 3 seconds. In total, 50 trials of this experiment were performed.

We will assume that the probability of spiking of the neuron in any 1 ms bin is independent of the trial, but changes with time. If we let $t = 1, \cdots, 3000$ be the index for the t^{th} time bin, the data $\mathbf{x}_t = (x_{t,1}, x_{t,2}, \cdots, x_{t,50})$ consist of the outcomes of a sequence of 50 independent Bernoulli trials, each with probability p_t.

We will compare filtered estimates of p_t obtained using a binary state-space model with those obtained using a moving average. For each $t = 1, \cdots, 3000$, because the 50 Bernoulli trials are independent, it is sufficient to record the total number $y_t = \sum_{k=1}^{50} x_{t,k}$ of spikes in the t^{th} time bin. The vector $\mathbf{y} = (y_1, y_2, \cdots, y_{3000})$ is provided to you in the file 'ThalamusStateSpace.p'.

(a) Plot the peri-stimulus time histogram (PSTH) of the data. The PSTH estimate of the probability of spiking in each time bin is computed as follows

$$p_t^{PSTH} = \frac{y_t}{50}, t = 1, \cdots, 3000. \tag{1}$$

What is the probabilistic model that underlies the PSTH?

(b) Let p_t^W be the moving-average estimate of the probability of spiking in the t^{th} window

$$p_t^W = \frac{\sum\limits_{k=t_0}^{t} y_k}{50 \times W}, \tag{2}$$

where W is the length of the window (e.g. $W = 10$) and $t_0 = \max(1, t - W + 1)$ accounts for the fact that, for $t < W$, there are fewer than W points to compute the moving average. Plot p_t^W for $W = 10, 50, 100, 200$. How does the smoothness of moving-average estimate depend on W?

(c) In the state-space framework, we compute the filtered estimate of p_t using a so-called binary-filter algorithm, which is a generalization of the Kalman filter for binary data such as those that come from a point-process. We denote by $p_{t|t}$ the filtered estimate of p_t obtained in the state-space framework. Given initial conditions $\theta_{0|0}$ and $\sigma^2_{0|0}$, and a parameter σ^2, we compute $p_{t|t}$ for $t = 1, \cdots, 3000$ using the following recursion

$$
\begin{aligned}
\theta_{t|t-1} &= \theta_{t-1|t-1} & (3)\\
\sigma^2_{t|t-1} &= \sigma^2_{t-1|t-1} + \sigma^2 & (4)\\
p_{t|t-1} &= \frac{\mathrm{e}^{\theta_{t|t-1}}}{1+\mathrm{e}^{\theta_{t|t-1}}} & (5)\\
\sigma^2_{t|t} &= \left(\frac{1}{\sigma^2_{t|t-1}} + T \cdot p_{t|t-1}(1-p_{t|t-1})\right)^{-1} & (6)\\
\theta_{t|t} &= \theta_{t|t-1} + \sigma^2_{t|t}(y_t - T \cdot p_{t|t-1}) & (7)\\
p_{t|t} &= \frac{\mathrm{e}^{\theta_{t|t}}}{1+\mathrm{e}^{\theta_{t|t}}}, & (8)
\end{aligned}
$$

where $T = 50$ for the thalamic data. Implement a the binary filtering algorithm. Plot $p_{t|t}$ for $\sigma^2 = 0.1, 0.01, 0.001, 0.0001$. Use $\theta_{0|0} = -3.5$ and $\sigma^2_{0|0} = 0$. How does the smoothness of estimate from the binary filter depend on σ^2?

You may find the '`lfilter()`' function from 'scipy.signal' helpful.

Problem 2

In this problem, we will use the binomial state-space model from to analyze data from the Twitter accounts of Hillary Clinton and Donald Trump from January 1st 2016 to February 1st 2017.

The data are stored in the pickle files each attached with a candidate's name. Each files consist of a binary matrix $Y_{k,t}$ of size 397×1440, where $Y_{k,t}$ is a Bernoulli random variable that indicates the presence or absence of a Tweet in the t^{th} minute of day k. There are a total of 397 days in the period considered and 1440 minutes in a 24 hour period. The notebook provided shows you how to load and plot the data.

From the matrix $Y_{k,t}$ we can form two vectors $\mathbf{y}_t^{WD} \in \{0,1\}^{1440}$ and $\mathbf{y}_k^{AD} \in \{0,1\}^397$, which represents respectively the Tweeting activity within a day (WD). and the Tweeting activity across days (AD).

(a) Each entry of $\mathbf{y}_t^{WD}$ is a Binomial(397,p_t) random variable. Use the binomial filter from the previous problem to compute $p_{t|t}, t = 1, \cdots, 1441$. Use $\theta_{0|0} = -4.5, sigma^2_{0|0} = 0$ and $\sigma^2_\epsilon = 0.15$. Label the x-axis of your plot as in the plots from the example notebook. The units of $p_{t|t}$ are in Tweets/minute; multiply the y-axis by 60 to obtain units in Tweets/hour.

(b) Each entry of $\mathbf{y}_k^{WD}$ is a Binomial(1440,p_k) random variable. Use the binomial filter from the previous problem to compute $p_{t|t}, t = 1, \cdots, 397$. Use $\theta_{0|0} = -4.5, \sigma^2_{0|0}$ and $\sigma^2_\epsilon = 0.4$. Label the x-axis of your plot as in the plots from the example notebook. The units of $p_{k|k}$ are in Tweets/minute; multiply the y-axis by 60 to obtain units in Tweets/hour.

Problem 3

In this problem this problem, we will study the steady-state behavior of the *discrete-time* Kalman filter/smoother and draw a parallel with kernel methods. In particular, we will show that posterior mean of the discrete-time Kalman smoother is closely related to a *continuous-time* function in a reproducing-kernel Hilbert space (RKHS) by an exponential kernel function.

Consider the following simple linear state-space model

$$\begin{cases} \theta_t = \theta_{t-1} + \epsilon_t, & \epsilon_t \sim N(0, \sigma_\epsilon^2) \text{ i.i.d.} \\ y_t = \theta_t + \delta_t, & \delta_t \sim N(0, \sigma_\delta^2) \text{ i.i.d.}, t = 1, \cdots, T. \\ \theta_0 = 0. \end{cases}$$

Given initial conditions $\theta_{0|0} = 0$ and $\sigma^2_{0|0} = 0$, the Kalman filter is given by the following recursive equations

$$\begin{aligned} \theta_{t|t-1} &= \theta_{t-1|t-1}, \\ \sigma^2_{t|t-1} &= \sigma^2_{t-1|t-1} + \sigma^2_\epsilon \\ \theta_{t|t} &= \theta_{t|t-1} + \frac{\sigma^2_{t|t-1}}{\sigma^2_\delta + \sigma^2_{t|t-1}}(y_t - \theta_{t|t-1}), \\ \sigma^2_{t|t} &= \frac{\sigma^2_\delta \sigma^2_{t|t-1}}{\sigma^2_\delta + \sigma^2_{t|t-1}}. \end{aligned}$$

We will concern ourselves with the so-called steady-state behavior of these equations, when we let $\lim_{t\to\infty} \sigma^2_{t|t-1} \to \sigma^2_\infty$–the so-called steady-state. The full set of filtering equations simplifies to

$$\begin{aligned} \theta_{t|t} &= \theta_{t-1|t-1} + c_\infty \left(y_t - \theta_{t-1|t-1}\right) \\ &= c_\infty y_t + (1 - c_\infty)\theta_{t-1|t-1}, \\ c_\infty &= \frac{\sigma^2_\infty}{\sigma^2_\delta + \sigma^2_\infty} \end{aligned}$$

(a) Given an initial condition $\theta_{0|0} = 0$, show that, for $t \geq 1$, $\theta_{t|t}$ can be expressed as follows

$$\theta_{t|t} = \alpha \sum_{k=1}^{t} \beta^{t-k} y_k, \tag{9}$$

for constants α and β that depend on c_∞.

(b) Let $\mathbf{y} = (y_1, \cdots, y_T)$, $H_t = \{y_{t-1}, \cdots, y_1\}$. By definition, $\theta_{t|t} = E[x_t|y_t, H_t]$. Let $\theta_{t|T} = E[\theta_t|\mathbf{y}]$, i.e. the posterior mean from the Kalman smoother. Argue intuitively (by symmetry, i.e. if you were to start the Kalman filter from the last time point $t = T$) from part (a) that

$$\theta_{t|T} = \alpha \sum_{k=1}^{T} \beta^{|t-k|} y_k, \tag{10}$$

and that if only $y_{t_1}, \cdots, y_{t_R}, R << T$ are nonzero (only R entries of $\mathbf{y}$ are nonzero, and their indices are $t_1, \cdots, t_R \subset \{1, \cdots, T\}$, this simplifies to

$$\theta_{t|T} = \alpha \sum_{r=1}^{R} \beta^{|t-t_r|} y_{t_r}. \tag{11}$$

(c) Let $t \in [0, T]$ and suppose $\theta(t)$ is a function that belongs to the RKHS $\mathcal{H}$ induced by the exponential kernel

$$\kappa(t_i, t_j) = \mathrm{e}^{-\gamma|t_i - t_j|}, \gamma > 0. \tag{12}$$

Suppose we are given noisy observations $y_i = f(t_i) + \epsilon_i$, $\epsilon_i \sim$ i.i.d. $\mathcal{N}(0, \sigma^2)$ of an unknown function $f(t)$, $t_i \in [0, T], i = 1, \cdots, n$. Consider the problem

$$\hat{\theta}(t) = \underset{\theta(t) \in \mathcal{H}}{\operatorname{argmin}} \frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \theta(t_i))^2 + \lambda \, ||\theta(t)||_{\mathcal{H}}^2 . \tag{13}$$

Express $\hat{\theta}(t)$ as a function of $\kappa(\cdot, \cdot)$, $t_i, i = 1, \cdots, n$ and an unknown set of n parameters $\alpha_i, i = 1, \cdots, n$.

(d) Compare $\hat{\theta}(t)$ to $\theta_{t|T}$ from part (b). Argue that that $\hat{\theta}(t)$ is a continuous-time counterpart of $\theta_{t|T}$.

N.B.: The advantage of the discrete-time formulation is its recursive nature, which yields tremendous savings in terms of memory and computation. Recall that to solve the problem of Equation 13, an $n \times n$ kernel matrix must be inverted!