Note:

1. Due 11:59PM, November 21, 2016.
2. Electronic submission to your instructor's email.
3. You are VERY MUCH encouraged to form teams to discuss proofs and program algorithms. If so, please acknowledge your teammate(s)' contributions at the beginning of your submitted homework. You must independently write your homework based on your own understanding.
4. Choose any programming language you like, R, Python, Matlab, C/C++, Julia, etc.

Examples and Implementations

[50 points; Problem Credit Peter Bartlett] Kalman Filter:
The data in file *kalman_filter.data* on the course web site contains noisy measurements of the location of a particle moving in the plane, subject to gravity, random forces, and drag. Each line of the file consists of the measurements $y_t \in R^2$ of the location at time *t = 1, . . . , T*. The true location of the particle at time t is $x_t \in R^2$, and its velocity is $\dot{x}_t \in R^2$. The equations of motion are

$$x_{t+1} = x_t + \dot{x}_t,$$
$$\dot{x}_{t+1} = 0.98\dot{x}_t - 0.02x_t + w_t,$$

where $w_t \sim N(0, 0.05I_2)$. The observations $y_t \in R^2$ have the form

$$y_t = x_t + v_t,$$

where $v_t \sim N(0, 100I_2)$. Suppose also that

$$\binom{x_1}{\dot{x}_1} \sim N(0, 5I_4).$$

(a) Plot the particle's true location x_t (from the file *kalman_filter.true* on the web site).
(b) Plot the observations y_t of the particle's position, on top of a plot of the true location.
(c) Explain how to estimate the particle's initial state (that is, its position x_1 and velocity $\dot{x}_1$ at the initial time t = 1) from the noisy measurements $y_1, \dots, y_T$.
(d) Calculate the maximum a posteriori probability initial state given the data in *kalman_filter.data*.
(e) For each t, plot the vector $\hat{x}_t$ of locations that maximize the probability $p(x_t \mid y_1, \dots, y_T)$ (we call this Maximum A Posteriori, or MAP estimate), on top of a plot of the true location. Include in the plot an arrow from x in the direction of the MAP initial velocity.

[Extra Credit; Theory Question; 25points] Suppose that we have a linear state space model with Gaussian disturbances, and we wish to estimate the initial state x_1 from the noisy observations $y_1, \ldots, y_T$, as in the previous question. It seems reasonable that as *T* increases, later observations provide less information about the initial state. In this question, we investigate this property.

Suppose that

$$
\begin{aligned}
&x_t \in R^p, y_t \in R^p, \\
x_{t+1} &= Ax_t + w_t, \qquad w_t \sim N(0, Q), \\
y_t &= Cx_t + v_t, \qquad v_t \sim N(0, \sigma^2)
\end{aligned}
$$

(Notice that $C \in R^{1\times p}$ is a row vector.)

(a) What is the conditional distribution of y_T given x_1?
(b) Suppose that the *C* vector has unit length and the matrix *A* is such that for all $v \in R^p$, $\|A^t v\| \leq \alpha^t \|v\|$, where $\alpha < 1$. Define the $T \times p$ matrix

$$
O_T = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{T-1} \end{bmatrix}
$$

(Matrices like O_T are called *observability matrices*: in a deterministic system, the rank of O_T characterizes whether the initial state can be inferred from subsequent observations.) Consider what happens when we start the system in two different initial states, x_1 and $\widetilde{x_1}$. Give an upper bound on the KL-divergence between $p(y_T \mid x_1)$ and $p(y_T \mid \widetilde{x_1})$ as a function of $\|x_1 - \widetilde{x_1}\|$, α, *T*, σ^2, and $tr(O_T Q O_T{}')$.

[50 points; Problem Credit Ciprian Crainiceanu] Download the Framingham data from the course website. Implement your own code in R for Bayesian inference based on MCMC simulations for the following Bernoulli model with covariate measurement error

$$
\begin{cases}
Y_i \sim Bernoulli(p_i) \\
logit\,(p_i) = \beta_0 + \beta_1 X_i + \beta_2 Z_i \\
W_{ij} \sim Normal(X_i, \sigma_U^2)
\end{cases}
$$

where Y_i is the CHD status, X_i is the true log SBP, W_{ij} are the two log SBP measurements, and Z_i is the smoking status for subject *i*.

a. Write the full conditionals and discuss the various options available for simulation.

b. Provide posterior means, standard deviations and credible intervals for model parameters.

c. Provide histories of the chains and discuss their quality with respect to length of burn-in periods and mixing properties.

d. Provide documented R code.

e. Compare your results with a naïve analysis that would simply replace X_i by $(W_{i1}+W_{i2})/2$.

f. Summarize your experience.