Data Assimilation and Climate
AIMS Rwanda, March, 2020

Assignment 1
Maximum Likelihood Estimation and Bayesian Estimation

For all problems consider the process model

$$X_{i+1} = \alpha X_i + \delta_i, \quad i = 1, 2, 3,$$

and data model

$$Y_i = X_i + \epsilon_i, \quad i = 2, 3.$$

1. Maximum Likelihood Estimation.
   Assume $\delta_i = 0$ and $\epsilon_i \sim \mathcal{N}(0, \tau^2)$ for all i . Then

   $$Y_2 \sim \mathcal{N}(\alpha x_1, 1 + \tau^2), \quad Y_3 \sim \mathcal{N}(\alpha^2 x_1, 1 + \tau^2 + \alpha^2), \quad \text{Cov}(Y_2, Y_3) = \alpha.$$

   The joint distribution of $\mathbf{Y} = (Y_2, Y_3)^T$ given X_1 is normal with density

   $$f_{Y|x_1} \propto \exp\left(-1/2(y_2 - \alpha x_1, y_3 - \alpha^2 x_1)\mathbf{\Sigma}_Y^{-1}(y_2 - \alpha x_1, y_3 - \alpha^2 x_1)^T\right).$$

   (a) Use the general matrix $\text{Cov}(\mathbf{Y})$ found in class to find the covariance matrix $\mathbf{\Sigma}_Y$.
   (b) Evaluate the exponent of the joint density $f_{Y|x_1}$. (Hint: The dimension of the exponent is 1).
   (c) Estimate x_1 from observations Y_2 and Y_3 by maximizing the density $f_{Y|X_1}$ with respect to x_1.
   (d) Discuss how much the data Y_2 and Y_3 effect the re-analysis if (i) $0 < \alpha < 1$ and (ii) $\tau \ll 1$.

2. Bayesian Estimation.
   Assume $X_1 \sim \mathcal{N}(\mu_0, \sigma^2)$ and $\epsilon_i \sim \mathcal{N}(0, \tau^2)$. Let $\mathbf{X} = (X_1, X_2, X_3, X_4)^T$ and $\mathbf{Y} = (Y_2, Y_3)^T$, then

   - *prior distribution* is $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{\Sigma}_X)$
   - conditional distribution of the data given process model is $\mathbf{Y}|\mathbf{X} \sim \mathcal{N}(\mathbf{H}\mathbf{X}, \mathbf{\Sigma}_Y)$
   - *posterior distribution* is $\mathbf{X}|\mathbf{Y} \sim \mathcal{N}(\boldsymbol{\mu}^*, \mathbf{\Sigma}^*)$

   Note that for this example $\mathbf{\Sigma}_Y$ is the same as in (**??**), $\boldsymbol{\mu}$ and $\mathbf{\Sigma}_X$ are in the Bayes lecture activity, while $\mathbf{H}$, $\boldsymbol{\mu}^*$ and $\mathbf{\Sigma}^*$ are given in the Bayes lecture notes.

   (a) Use the sympy package in Python to show that the variance of the *posterior distribution* for X_1 (i.e. reanalysis) is

   $$\text{var}(X_1|\mathbf{y}) = \frac{\sigma^2(\tau^4 + \tau^2(\alpha^2 + 2) + 1)}{\tau^4 + \tau^2((\alpha^2 + 1)\alpha^2\sigma^2 + \alpha^2 + 2) + \alpha^2\sigma^2 + 1}.$$

   Show that

   $$\text{var}(X_1|\mathbf{y}) < \text{var}(X_1)$$

   and discuss the relationship between the posterior variance and the prior variance.

(b) Use the sympy package in Python to show that the variance of the *posterior distribution* for X_3 (i.e. filtering) is

$$\mathrm{var}(X_3|\mathbf{y}) = \frac{\alpha^4\sigma^2(\tau^4 + \tau^2(\alpha^2+2)+1)}{\tau^4 + \tau^2((\alpha^2+1)\alpha^2\sigma^2 + \alpha^2 + 2) + \alpha^2\sigma^2 + 1}.$$

Show that

$$\mathrm{var}(X_3|\mathbf{y}) < \mathrm{var}(X_3)$$

and discuss the relationship between the posterior variance and the prior variance.