Data Assimilation and Climate

AIMS Rwanda, March, 2020

Lecture Notes

Bayesian approach to Data Assimilation

Covariance

A measure of how much two random variables vary together

$$\mathrm{Cov}(X) = E[(X-\mu_x)(Y-\mu_y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_x)(y-\mu_y)f_{XY}(x,y)\ dxdy$$

Consider vectors $\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{pmatrix}$ and $\boldsymbol{\mu} = \begin{pmatrix} \mu_{x_1} \\ \mu_{x_2} \\ \vdots \\ \mu_{x_N} \end{pmatrix}$

$$\mathrm{Cov}(\mathbf{X}) = E[(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})^T] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})^T f_X(x_1,\dots,x_N)dx_1\dots dx_N$$

$$= \begin{bmatrix} \mathrm{Var}(X_1) & \mathrm{Cov}(X_1,X_2) & \dots & \dots & \mathrm{Cov}(X_1,X_N) \\ \mathrm{Cov}(X_1,X_2) & \mathrm{Var}(X_2) & \mathrm{Cov}(X_2,X_3) & \dots & \mathrm{Cov}(X_2,X_N) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(X_N,X_1) & \mathrm{Cov}(X_N,X_2) & \dots & \dots & \mathrm{Var}(X_N) \end{bmatrix}$$

Lecture Activities 1(a)(b)(c)

Useful formulas

$$\begin{aligned} E[X+c] &= E[X]+c \\ E[cX] &= cE[X] \\ E[X+Y] &= E[X]+E[Y] \end{aligned}$$

$$\begin{aligned} \mathrm{Var}(c) &= 0 \\ \mathrm{Var}(X+c) &= \mathrm{Var}(X) \\ \mathrm{Var}(cX) &= c^2\mathrm{Var}(X) \\ \mathrm{Var}(X+Y) &= \mathrm{Var}(X)+2\mathrm{Cov}(X,Y)+\mathrm{Var}(Y) \end{aligned}$$

$$\begin{aligned} \mathrm{Cov}(X,c) &= 0 \\ \mathrm{Cov}(X+c,Y+k) &= \mathrm{Cov}(X,Y) \\ \mathrm{Cov}(X+Y,Z) &= \mathrm{Cov}(X,Z)+\mathrm{Cov}(Y,Z) \end{aligned}$$

Lecture Activities 1(d)

Bayesian Approach to Data Assimilation

- There is a *prior distribution* on $\mathbf{X} = (X_1, X_2, \ldots, X_N)^T$.

Example

If $X_2 \sim \mathcal{N}(\alpha\mu, \alpha^2\sigma^2)$ the *distribution* is

$$f_{X_2}(x_2; \alpha, \mu, \sigma) = \frac{1}{\alpha\sigma\sqrt{2\pi}} e^{-1/2(x_2-\alpha\mu)^2/\alpha^2\sigma^2}$$

Example

If $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{\Sigma}_X)$ the *distribution* is

$$f_{\mathbf{X}}(\mathbf{x}; \boldsymbol{\mu}, \mathbf{\Sigma}_X) \propto e^{-1/2(\mathbf{x}-\boldsymbol{\mu})^T \mathbf{\Sigma}_X^{-1}(\mathbf{x}-\boldsymbol{\mu})}$$

where $(\mathbf{x} - \boldsymbol{\mu})^T \mathbf{\Sigma}_X^{-1} (\mathbf{x} - \boldsymbol{\mu})$

$$= (x_1 - \mu_1, x_2 - \mu_2, \ldots, x_N - \mu_N) \left(\qquad\qquad\qquad \right)^{-1} \begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \\ \vdots \\ x_N - \mu_N \end{pmatrix} \in \mathbb{R}$$

Bayesian Approach to Data Assimilation (continued)

- There is a conditional distribution on the data $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_N)^T$.

Example

If $\epsilon_2 = y_2 - x_2 \sim \mathcal{N}(0, \tau^2)$ the distribution is

$$f_{Y_2|\mathbf{x}}(y_2|x_1, x_2; \tau) = \frac{1}{\tau\sqrt{2\pi}} e^{-1/2(y_2-x_2)^2/\tau^2}$$

Example (Assignment 1)

If $Y_2 \sim \mathcal{N}(\alpha x_1, 1 + \tau^2)$ and $Y_3 \sim \mathcal{N}(\alpha^2 x_1, 1 + \tau^2 + \alpha^2)$ the distribution is

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

- Given the *prior distribution* and conditional distribution on the data, the goal is to find the *posterior distribution* on X, $f_{X|y}(\mathbf{x})$. The mean of $f_{X|y}(\mathbf{x})$ gives our state estimates $\mathbf{x}$ that assimilate the data $\mathbf{y}$. The variance of $f_{X|y}(\mathbf{x})$ gives uncertainty in the state estimates.

Bayes Theorem

$$f_{X|y} = \frac{f_{Y|x}(\mathbf{y})f_X(\mathbf{x})}{f_Y(\mathbf{y})}$$

The distribution on the data, $f_Y(\mathbf{y})$, is unknown. However, we don't need to calculate it because it is constant with respect to the unknowns $\mathbf{x}$.

$$f_{X|y} \propto f_{Y|x}(\mathbf{y})f_X(\mathbf{x})$$

Before we do an example consider we are working with vectors $\mathbf{X} = (X_1, X_2, \ldots, X_N)$ and $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_N)$ and we no longer wish to write expressions and equations element wise. Recall our example with $\{X_1, X_2, X_3, X_4\}$ and $\{Y_1, Y_2\}$, we can re-write this in vector-matrix notation as

$$\begin{pmatrix} Y_2 \\ Y_3 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{pmatrix} + \begin{pmatrix} \epsilon_2 \\ \epsilon_3 \end{pmatrix}$$

This gives data model $\mathbf{Y} = \mathbf{HX} + \epsilon$.

Example of Bayesian Data Assimilation

- Given *prior distribution* $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}_X)$
- Given conditional data distribution $\mathbf{Y}|\mathbf{X} \sim \mathcal{N}(\mathbf{HX}, \boldsymbol{\Sigma}_y)$
- Bayes Rule tells us

$$\begin{aligned} f_{X|y} &\propto f_{Y|x} f_X \\ &\propto e^{-1/2(\mathbf{y}-\mathbf{Hx})^T \boldsymbol{\Sigma}_Y^{-1}(\mathbf{y}-\mathbf{Hx})} e^{-1/2(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}_X^{-1}(\mathbf{x}-\boldsymbol{\mu})} \\ &= \exp\left(-1/2((\mathbf{y}-\mathbf{Hx})^T \boldsymbol{\Sigma}_Y^{-1}(\mathbf{y}-\mathbf{Hx}) + (\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}_X^{-1}(\mathbf{x}-\boldsymbol{\mu}))\right) \end{aligned}$$

$f_{X|y}$ also follows a normal distribution and we use completing the square (leaving out details) to get it in the form

$$f_{X|y} \propto e^{-1/2(\mathbf{x}-\boldsymbol{\mu}^*)^T (\boldsymbol{\Sigma}^*)^{-1}(\mathbf{x}-\boldsymbol{\mu}^*)}$$

where

$$\begin{aligned} \boldsymbol{\mu}^* &= \boldsymbol{\mu} + \mathbf{K}(\mathbf{y} - \mathbf{H}\boldsymbol{\mu}) \\ \boldsymbol{\Sigma}^* &= (\mathbf{I} - \mathbf{KH})\boldsymbol{\Sigma}_x \end{aligned}$$

with $\mathbf{K} = \boldsymbol{\Sigma}_x \mathbf{H}^T (\boldsymbol{\Sigma}_y + \mathbf{H}\boldsymbol{\Sigma}_x \mathbf{H}^T)^{-1}$ the gain matrix.

Lecture Activities 3.